ANALYSIS OF THE USEFULNESS OF AUTOMATICALLY PROCESSED ERTS MULTISPECTRAL DATA FOR GEOLOGIC PURPOSES IN GEORGIA

A THESIS

Presented To

The Faculty of the Division of Graduate Studies

by

Nickolas L. Faust

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Geophysical Sciences

Georgia Institute of Technology

January, 1976

Analysis of the Usefulness of

Automatically Processed ERTS Multispectral

Data for Geologic Purposes in Georgia

Approved:

Dr. L. T. Long, Chairman

Date Approved by Chairman: 1-21-76

ACKNOWLEDGMENTS

The author would like to gratefully acknowledge the direction and understanding of Dr. L. T. Long in the writing of this thesis. He would also like to thank his wife, Cheri, for her typing assistance, unending patience, and support, without which this thesis could not have been completed. This work was partially supported under NSF Grant #Ga 12391 and NASA Contract #NA58-30653. Computer time costs were partially provided by the School of Geophysical Sciences and the Engineering Experiment Station at Georgia Tech.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Cont.)

LIST OF FIGURES (Cont.)

SUMMARY

This thesis presents an evaluation of the application of computer-implemented pattern-analysis techniques to Earth Resources Technology Satellite (ERTS) data as a source of geological information in the State of Georgia. The dense vegetative cover in the southeastern United States prevents direct classification of rock units. However, the interpretation of the spatial distribution of the vegetation classes generated from ERTS data was found to give an insight as to the location of geologic structure in Georgia.

The digital analysis techniques employed in this study included clustering, supervised classification, linear classification, and Fast Fourier Transformation (FFT) analysis. Clustering was used to delineate on a map the location of groups of ERTS resolution elements that appeared similar in the four spectral bands of ERTS data. Once clusters were produced and displayed in map form, the clusters were identified as representing vegetation groups, water, or groups of manmade objects by visual comparison with ground truth information such as maps and low altitude photographs. The statistics of each identified cluster could then be calculated. A supervised classification system was then used which classified each ERTS resolution element by comparing its spectral response to the spectral response for previously defined statistical classes.

Supervised classification was found to be useful in deliniating a vegetation group that could be correlated directly to a particular

soil group. This soil group was then traced to a major rock unit associated with the Brevard Fault. A similar technique was found to be useful in the detection of ultrabasic rocks in northeast Georgia.

A quantization of the spatial extent of ERTS data is studied by using an FFT algorithm. A measure of the spatial structure about a particular resolution element was calculated by computing integrals in various angular and radial regions centered on the resolution element. A spatial classifier was then designed which utilizes the FFT integrals as additional data for classification of ERTS data. By varying the way in which the FFT data were summed, features such as lineations and short and long wavelength regions can be enhanced. FFTs were computed on several lines of ERTS data over Atlanta. Plots of the FFT integral magnitudes show the ability of the FFT technique to describe spatial structure.

CHAPTER I

INTRODUCTION

Remote Sensing in its purest (or broadest) form refers to the gathering of information from a distance. However, since the late 60's and early 70's the term remote sensing has been commonly used in a narrower sense to refer to the interpretation of photographic or digital data obtained from space or high altitudes. The large research and publicity efforts which have dominated the remote sensing field are a result of the massive funding of earth resource inventory projects by the National Aeronautics and Space Administration (NASA) in efforts to utilize the new types of data that became available as a result of the various space programs. Remote sensing of the earth's resources has become a popular subject. The realization that the raw materials of the earth are limited further stimulated interest in making an inventory of the world's resources.

Information which can be extracted from remote sensing data could be valuable to many federal, state, and local government agencies as well as individuals and private companies who need a comprehensive assessment of the resources present in some section of the earth's surface. For example, foresters may wish to know the extent of pine forests in California while an urban planner may wish to be able to determine the best location for a new airport. The analysis of remote sensing data makes available quantitative information which can be applied to such problems.

Geologists have long been users of such remote sensing data as aerial photographs. These pictures often aid geologists in the determination of the boundaries between geological units. As higher altitude aerial photographs became available from NASA, geologists were, with the help of minimal ground-gathered data, able to map larger and larger regions because of the perspective offered by the photographs. Similarly, satellite photographs from the Earth Resources Technology Satellite (ERTS) shown in Figure 1 and SKYLAB have proven to be valuable tools to geologists concerned with regional geology. The perspective given by the great height of the satellite and the resulting ability to see large regions of the earth in one picture frame often enabled geologists to detect regional trends such as fault traces which were not obvious on lower level photography.

In this study several state-of-the-art digital analysis techniques are applied to ERTS data taken over Georgia. The main objective is to evaluate the ability of these techniques to provide information that can be related to the geology of Georgia.

CHAPTER II

BACKGROUND

To date, by far the majority of all digital processing of remote sensing multi-spectral scanner (MSS) data has been done by Purdue University's Laboratory for Agricultural Remote Sensing (LARS) and the University of Michigan's Environmental Research Institute of Michigan (ERIM). Similar processing techniques are now being developed by institutions throughout the United States including the Engineering Experiment Station of Georgia Tech. However, notably few examples of automatic processing techniques applied to geologic problems have come from such heavily vegetated regions as the southeastern United States. This investigation represents an effort to apply automatic processing techniques in an area where the geologic formations are shielded by vegetative cover.

Research in the digital processing of remote sensing data can be separated into two areas: development of techniques, and application of techniques. Y. C. Ho and A. Agrawala (1968) and G. Nagy (1968) gave summaries of the state-of-the-art analysis techniques applicable to remote sensing by 1968. Predictions of the massive amounts of data to be returned by earth resources sensors prompted Haralick (1968) and Langrebe (1969) to realize the necessity for automatic processing of earth resources data which would eventually consume significant computer time.

The data commonly cluster into distinct spectral signatures which

can be identified with specific classes of sources. Clustering is an automatic technique for identifying concentrations (i.e. clusters) of spectral signatures. Clustering requires no previous information as to the probable clusters present in an area for which MSS data are available. Su, Jayroe, and Cummings (1972) refined clustering by incorporating a boundary detection and tracing algorithm. This algorithm was found to increase the speed and precision by classifying whole areas within a closed boundary as being from the same class without testing each element within the boundary. This was an early attempt to use spatial information to improve classification when compared to ground surveys.

In agriculture the precision of the classification of spectral data associated with specific crops is often determined by comparison to manually gathered estimates of crop area. For example, an agriculture expert may walk around a field and satisfy himself that the only crop planted in the area is wheat. Thus he calculates the area of wheat as the area of that field even though there may be bare spots in the unexamined portion of the field.

Crane and Richardson (1972) and Bond and Atkinson (1972) increased the speed of the operations required to associate spectral signatures with specific classes. They accomplished this by using a linear classifier which considers the MSS data one channel at a time during analysis. In their opinion, degradation of accuracy in classification due to the use of the linear classifier was more than compensated for by the speed of the system.

Gramenopoulos (1974) attempted to use spatial information in a classification of ERTS data by taking the Fast Fourier Transform (FFT)

of a 32 by 32 block of ERTS resolution elements. Integrals of regions of the spectral amplitudes in the wave number domain were used to define a data vector representing the 32 by 32 block. The entire area being considered was partitioned into similar blocks and the resulting data vectors were subjected to cluster analysis. This method proved to be effective in the separation of gross terrain types such as mountains and plains, because the spectral regions integrated can be made sensitive to the size or trend of a pattern.

In the applications area, Weston (1973) and Baumgardner, et al. (1973) showed that computer processed ERTS data could be used for soil mapping and vegetation analysis. Baumgardner, et al. (1973) point out that water detection, gross soil pattern analysis, and detection of damaged crops are possible with ERTS data.

A landuse mapping system using clustering and map scale printouts of clustered digital data was presented by Ellefsen, et al. (1973). The ability to scale the ERTS output was a significant contribution since the results could be compared directly to maps or photographs to check on the accuracy of automatic classification. A similar system has been developed at Georgia Tech (Spann and Faust, 1974).

In 1972 the digital techniques developed in previous years began to be applied to geological and other problems related to earth resources. Smedes (1972) used clustering techniques on 12 channel MSS data to obtain a rock-recognition map of Yellowstone National Park. Since rock outcrops are numerous in this area, direct spectral classification of rock types was achieved with greater than 90 percent accuracy.

Saunders, et al. (1974) showed that ERTS data could be utilized

in oil exploration. They emphasized the visual interpretation of regional structural trends. Digital processing was investigated, but was not found to be advantageous because of the excessive computer time involved in the classification of an ERTS scene and the lack of available classification algorithms which can incorporate the spatial extent of classes.

CHAPTER III

DATA BASE

The quantity measured by the ERTS MSS is the intensity of the electromagnetic energy emitted from a source at the earth's surface. If an energy flux, ϕ, is emitted from a source, its intensity measured at an MSS at a distance R from the source is given by

$$I = \phi/R^2$$

The energy flux from the source in the visible spectrum results from the reflection of solar radiation. Thus, if energy with an incident flux, ϕ_i, strikes a surface with a reflection coefficient of ρ the total outward energy flux from each unit area is $(\rho\phi_i)$. (Figure 2). The rest of the energy is either transmitted or absorbed by the surface. A simple model for the intensity measured by the ERTS MSS is

$$S = ((\rho\phi_I\ T)/R^2) + b$$

where S is the signal at the MSS

ρ is the surface reflectance

T is the atmospheric transmittance

b is the background radiation entering the imaging mirror

and R is the distance from the source to the imaging mirror.

Other factors influencing the signal are wavelength dependence, angle from normal to the imaging mirror, time of day, Raleigh and Mie scattering, and cloud cover effects.

Even though these other factors are present, we may still gain valuable insight into the characteristics of the reflecting ground element. For example, provided that cloud cover can be neglected we can expect the effects from the atmosphere and sun angle to be relatively constant. Thus, as long as the data being processed were obtained within a relatively short time period, variations in the signal intensity at the sensor will be due primarily to variations in the reflectance of the ground or its cover.

The sensor system on the ERTS satellite is a multispectral scanner (MSS). A MSS system collects radiance information in discrete bands of the electromagnetic spectrum. In the visible region of the spectrum it gathers the same information as a filtered black and white photograph, but in other spectral regions it obtains information that will be dependent on the response of the particular detector system. Several aircraft MSS systems are also now in use. The most widely publicized is the University of Michigan's twelve channel MSS system. For this thesis, a channel is defined as a band of the electromagnetic spectrum which is sensed and recorded as a single intensity value.

Information concerning the radiance intensity determined by any given sensor or scanner is normally stored in a digital form. An upper and lower limit on the radiance intensities is provided by calibrated sources for each scanner channel. The radiance measured between these

limits is then quantatized into one of 256 levels for each spot on the ground from which the scanner receives information. A given MSS channel stores the intensity datum as an 8 bit word.

Generally, a multispectral scanner consists of a rotating mirror system, a prism or beam splitter, and a bank of detectors. The scanner mirror rotates orthogonally to the direction of motion as shown in Figure 3. As the mirror passes through its arc it reflects incoming light from individual ground elements into the scanner interior. This energy passes through a prism and the different wavelengths are directed toward appropriate detectors. The size of the ground area imaged on the mirror at any one time determines the resolution of the instrument. Each image is called a resolution element. The size of a resolution element is directly dependent on both the altitude and the speed of the aircraft or spacecraft. For example, for low flying aircraft the resolution element size often lies between one and ten meters square, whereas for a satellite in orbit 1000 km above the earth the resolution element is on the order of 100 m square.

The ERTS satellite was inserted into a 1056 km circular orbit at an azimuth of 190 degrees. The primary remote sensing equipment aboard ERTS is the RBV television imaging system, the multispectral scanner, tape recording equipment, and a data transmission facility. The RBV camera system is designed to take photos of an area beneath the spacecraft 185.2 km square and provides an image on a cathode ray tube screen in the spacecraft. This image is scanned and digitized on board and transmitted to earth. The camera system uses filters to provide images in three spectral ranges; (1) .475-.575 μm, (2) .580-.680 μm (3) .698-

.830 μm. Unfortunately, the magnetic tape backup system for the RBV cameras failed soon after launch and the RBV system was shut down.

The multispectral scanner system aboard ERTS covers much the same spectral range as the RBV cameras. Unlike the RBV cameras, the radiances for each resolution element are measured and immediately recorded as digital information. The MSS system was originally regarded as a backup system for the RBV cameras, but data transmitted back to earth from it proved equal in quality to that from the RBV cameras. The MSS spectral bands, or channels, are (4) .4-.5 μm, (5) .6-.7 μm, (6) .7-.8 μm, (7) .8-1.1 μm, or, green, red, and two regions in the near infrared.

A typical element resolved by the ERTS MSS is 79 meters by 57 meters, or approximately 0.45 hectare. This resolution means that the signal sources may not be located any closer than $\pm$ 67 meters. For many uses this resolution is far too coarse, and ERTS data are not applicable. However, for studies of regional scope such as are useful in geological studies, ERTS may provide a regional mapping tool. This is the subject addressed by this thesis.

The data that are transmitted back to earth from ERTS are the four 8 bit words representing each of the four channels for each resolution element. A 4.5 hectare area of land would be represented 40 numbers in the range of 0 to 256.

CHAPTER IV

CLUSTERING

A clustering technique is a method of locating concentrations of data vectors in N dimensions. Given a data base which may be represented by an N dimensional density function,

$$D\ (X_1,\ X_2,\ \ldots,\ X_N),$$

the objective of clustering is to locate the position in N space of the points of maximum density. A sequential method of defining clusters is used in this thesis to estimate the positions of the maximum density. In general, the distance between two location vectors $\overline{A}$ and $\overline{B}$ in N space may be calculated by

$$d = \left(\sum_{i=1}^{N} (a_i - b_i)^2 \right)^{1/2}$$

where a and b are elements of the location vectors $\overline{A}$ and $\overline{B}$. For a clustering analysis the vector $\overline{B}$ is replaced by a mean vector of a cluster. Assuming that there are m vectors in the cluster, $\overline{V}_i$ with element v_{ik} the mean M can be calculated by

$$\bar{\bar{M}} = \begin{matrix} M_1 \\ M_2 \\ \vdots \\ M_m \end{matrix}$$

$$M_k = \frac{1}{m} \sum_{i=1}^{m} v_{ik}$$

where k represents the spectral channel.

If the vector $\bar{A}$ is within a constant prespecified spectral radius (Rsr) in N dimensions from the mean of a cluster, i.e. if

$$d = \left(\sum_{i=1}^{N} (a_i - M_i)^2\right)^{1/2} \leq Rsr$$

where a_i is the value of one element of the N dimensional data vector then the data vector is assigned to that cluster.

For an ERTS cluster analysis the data vectors are four-dimensional vectors made up of spectral measurements made by the MSS system. These four measurements are associated with one ERTS resolution element and are the intensities of light received by the detectors on the spacecraft in each of the four spectral bands.

Now consider the situation in Figure 3. The multispectral scanner scans a region normal to the flight path of the spacecraft. At any instant in time the rotating mirror displays an image or resolution

element representing 0.45 hectares on the ground and measurements in 4 regions of the spectrum are taken. The spacecraft velocity and the scanner rotation speed are such that after one scan line of data is taken, the spacecraft has moved forward enough so that the next scan line is contiguous to the first.

ERTS MSS data can be analyzed digitally using clustering (Wylie et al., 1974). The radiance values for the resolution elements of a particular area are represented in four space. An illustrative situation in three space is shown in Figure 4. The data points are distributed among several groupings which probably represent radiance values from the same or similar objects. For example, Group A might be radiance values from trees, Group B from buildings, and Group C from water. Using clustering, a description of each cluster by a mean vector and a chosen spectral radius is developed (Figure 5). If a resolution element vector falls within a predefined spectral radius of Group A mean, the resolution element is assigned to Group A. This follows similarly with other groups. If a vector does not fall within the prescribed spectral radius of any of the previously defined clusters, a new cluster is generated using that vector as the mean. This method may be biased due to the starting point in the data set since MSS data are usually considered sequentially. Another disadvantage to the sequential treatment is that if the cluster radii are chosen too small only a few points are allowed in a cluster. Thus, many additional clusters will be formed. The selection of cluster spectral radius values is essentially a trial and error procedure. As the number of clusters increases so does computer time and storage. This limits the

number of clusters that may be considered. In the analysis reported on in this thesis up to 20 clusters were considered. To eliminate the bias of the sequential building of clusters, a second iteration is made with each cluster generated in the first pass being weighted by the number of vectors it contained in the first pass. This has the effect of not allowing the means to change substantially in the second iteration. These clusters may often be identified with different objects on the ground such as water, rock, etc. Experience has shown that the choice of an appropriate radius in spectral space depends strongly on the area of study. The clustering results can be presented in map form by associating a character with one of the previously determined clusters. With such a map one can see the spatial location of the yet unidentified clusters. With ground truth information such as maps and aerial photographs, these clusters may be identified with classes representing major housing and development trends within a city as well as many other land uses.

CHAPTER V

SUPERVISED CLASSIFICATION

In the previous section a technique was described for locating concentrations or "clusters" of ERTS data vectors. However, a mean and a spectral radius in four dimensions may not adequately describe the distribution of data vectors derived from a particular material on the earth's surface. A more complex distribution may be described by a mean vector and unequal spectral radii in each dimension. The data are assumed to be normally distributed in each dimension.

Using clusters generated in the above section and subsequently identified as to source by use of ground truth information, a more accurate description of the distribution of the data vectors in a cluster may be achieved by considering the vectors enclosed in a particular cluster and calculating the density distribution of the data in each spectral dimension. In this process each cluster is tested to assure that the data values in each channel for all vectors in the cluster approximate a normal distribution. If the data are not found to be unimodal the cluster may not be considered appropriate for supervised classification. Figure 6 shows a bimodal distribution in one dimension. A normal distribution would be represented by an elipsoidal density function about the mean vector (Figure 7). Thus, if we wanted to describe an ellipsoid in 4 space we would need to calculate the mean and the direction and length of the semiaxes. This may be done by the

calculation of the variance of the data from the mean. The variance denoted by σ^2_n (n = 1, 4) is a measure of the elongation of the data in a particular direction. For J data vectors,

$$\sigma^2_n = 1/J \sum_{n=1}^{J} (a_{in} - u_n)^2$$

where n refers to the spectral channel.

In 95% of the cases considered, a randomly chosen data value "a" will fall in the region defined by $|a-u| \leq 2$ where u is the mean value and where a belongs to a normal distribution of values. In figure 8 σ is the standard deviation of one channel (i.e. dimension) in spectral response. This may be extended to N channels of data by considering that there is a standard deviation with each channel of data. If we assume that we are dealing with data distributed according to a normal distribution, we can only estimate the values for the mean and the standard deviation associated with a particular class. The variance of the data is equivalent to σ^2. In general, if many samples are available to calculate the mean value, the mean will approximate the true mean. If only a few resolution elements are considered there may be an error in the calculation of the mean for a particular class. A general measure of the precision of the mean calculation is given by Hays and Winkler (1971). In multivariate analysis, the variances in each of the spectral regions are not the only considerations. If data values in some channels depend on data values in other channels, the covariance may be represented in an N by N matrix. If there is not interdependence,

the channels are said to be independent and the covariance, zero. The mean, variance, and covariance for J data vectors are defined below.

$$\hat{M}_n = \frac{1}{J} \sum_{k=1}^{J} a_{kn}$$

$$\sigma^2_{nn} = \frac{1}{J} \sum_{k=1}^{J} (a_{kn} - u_n)^2$$

$$\sigma_{In}^2 = \frac{1}{J} \sum_{k=1}^{J} (a_{kI} - U_I)(a_{kn} - U_n)$$

and where a_{kn} is the n^{th} element of the k^{th} vector.

The square root of the covariances, σ_{IJ}, are correlation coefficients.

If a sufficient number of samples are used to define the above population, the diagonal elements of the covariance matrix will be the variances for each channel and the off-diagonal elements describe the interaction between channels of data, the covariances. A sample covariance matrix $\hat{\Sigma}$ is shown on the following page.

$$\hat{\Sigma} = \begin{matrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 & \sigma_{24}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 & \sigma_{34}^2 \\ \sigma_{41}^2 & \sigma_{42}^2 & \sigma_{43}^2 & \sigma_{44}^2 \end{matrix}$$

If the channels of data were independent then

$$\hat{\Sigma} = \begin{matrix} \sigma_{11}^2 & 0 & 0 & 0 \\ 0 & \sigma_{22}^2 & 0 & 0 \\ 0 & 0 & \sigma_{33}^2 & 0 \\ 0 & 0 & 0 & \sigma_{44}^2 \end{matrix}$$

Thus, given a sufficient number of resolution elements in a cluster that are identifiable with a class of material on the earth's surface, an estimated mean and a covariance may be computed for the total population of that material. By comparing each resolution element data vector to these estimates for each class of material we may decide if that data vector falls within a certain class of material, i.e., water.

Discriminant functions are developed in classification theory for special distributions of data. These discriminant functions are

the criteria by which a resolution element may be assigned to a particular class.

Assuming a normal distribution, the discriminant function g_i for a radiance vector $\bar{A}$ to be in the i^{th} class is in matrix notation.

$$g_i(\bar{A}) = -1/2\ (\bar{A} - \bar{M}_i)^t\ \hat{\Sigma}_i^{-1}\ (\bar{A} - \bar{M}_i) - d/2 \log 2\pi - 1/2 \log |\hat{\Sigma}_i| + \log P_i$$

where $\bar{M}_i$ is the mean vector for the i^{th} class $|\hat{\Sigma}_i|$ is the determinant of the covariance matrix, and $\hat{\Sigma}_i^{-1}$ is the inverse of the i^{th} class covariance matrix (Duda and Hart, 1973). In general the $d/2 \log 2\pi$ term is not a function of which class is considered. Assuming equal a priori probabilities P_i that a vector will fall into a certain class, the log P_i term is also constant. Thus, these terms may be ignored since we are comparing several computations all with the same constant terms. Hence a new discriminant function Q_i equivalent to the exponential of $g_i(\bar{A})$ can be given by

$$Q_i = \exp(g_i(\bar{A})) = \frac{e^{-1/2\ (\bar{A} - \bar{M}_i)^t\ \Sigma_i^{-1}\ (\bar{A} - \bar{M}_i)}}{|\hat{\Sigma}_i|^{1/2}}$$

Now for every radiance vector $\bar{A}$ a Q is calculated for each class previously defined. The vector is then assigned to the class that has the largest value of the discriminant function Q. This proceeds until all the resolution element vectors for the imaged area are processed. One pitfall of this method is that a vector is always assigned to one of

the classes even though it actually may not be similar to any of the classes. This pitfall may be eliminated by defining a threshold. If all classes have a Q value less than a threshold Q_t then the resolution element is not assigned to any class.

Since the $\hat{\Sigma}_i^{1/2}$ and the $\hat{\Sigma}_i^{-1}$ need only be calculated once for each class, the most time consuming part of the calculation is the quadratic computation of $(\overline{A} - \overline{M}_i)^t \hat{\Sigma}_i^{-1} (\overline{A} - \overline{M}_i)$. This quadratic computation must be calculated for each data vector, for each class consiered, and for each data dimension. There are approximately 7.5 million data vectors in one ERTS scene, and there are four ERTS channels. For ten classes then, the above computation must be performed 75 million times to process one ERTS scene. This is the reason that digital processing of ERTS data is often prohibitably expensive.

Thus, to summarize, the supervised method of classification uses distribution parameters assumed to fit a large number of samples to describe each class of data to which a resolution element vector may be assigned. Once these distribution parameters are calculated, the discriminant function must be calculated for each class for every data vector. The vector is then assigned to one of those classes by inspection of the discriminant functions.

CHAPTER VI

INTERPRETATION PROBLEMS

In evaluating data such as that obtained from ERTS, many unique interpretation problems arise. Two problems that persisted throughout the course of the work that led to this thesis were, first, the shielding of geologic structure by vegetation, and second, the inconvenience of a line printer output of digitally classified ERTS data.

In the southeast vegetative cover usually obscures the rocks and soils. Since classification of spectral signature returns is constrained by the source of the radiation, and since in the southeast the spectral signatures are dominated by the radiation from the vegetative cover, information about the geologic structures and soil types must be obtained by indirect means. One strategy is as follows. Initally, the ERTS data base over the considered area is subjected to a cluster analysis. By an appropriate choice of the cluster spectral radii a differentiation between certain vegetation types is obtained. Furthermore, many other clusters may be formed, some of which often represent classes which can be identified as something other than vegetation. These clusters may give some insight into the isolation of a class of spectral signatures related to the geologic structure.

Sometimes geologic formations which do not exhibit a unique spectral signature may be identified by certain distinctive patterns which appear when the clusters are plotted in map form. For example,

formations in an area may often be revealed by the presence of elongated vegetative patterns. Circular features which may indicate domal uplift may also be detected in the clustering because of the symmetrical distribution of vegetative types about the suspected uplift area. Thus, through clustering, vegetative patterns associated with geologic structure may be so enhanced that geologic structures are recognizable. These indirect interpretation techniques are in contrast to the direct employment of spectral signatures to identify prevalent rock types. Such direct identification is sometimes possible in the western United States where vegetative cover is slight, but it is rarely usable in the heavily vegetated eastern United States.

Since hardware limitations prevent the printing of computerized data at an arbitrary scale, a method was devised to rescale the raw resolution element data. The ERTS data are interpolated so that the size of an individual resolution element is changed and the resultant classification of the new data can be printed at the new scale. This scale change allows the results of a classification to be directly overlain on any scale map such as 7 1/2 minute U. S. Geological Survey quadrangle. Then, identification of various features is simplified.

CHAPTER VII

TEST SITE RESULTS

The results of this study apply to the southeastern United States in general. However, because of the availability of only a limited amount of digital ERTS data and the cost of computer processing, data from a few test sites (Figure 9) in Georgia are used. Initially, data from the Stone Mountain area, from the Cumberland Mall area in northwest Atlanta, from the Macon area, and from the area around Carters, Georgia were analyzed. Later, studies are extended to the Douglasville area and the area around Lake Chatuge.

Clustering analysis was applied to all four of the initial test sites. The results were mapped at a scale of 1:24000 with each symbol on the computer map representing a single cluster. Several preliminary tests were made to determine reasonable spectral radii for clustering before the results described in the examples below were obtained.

Cumberland Mall Area

Figure 10 shows the clustering results obtained for the Cumberland Mall test site. Comparing Figure 10 with Figure 11, an aerial photograph of the same area, one readily notes that the characters which represent distinct clusters may be correlated to land use in the area. For example, the cluster identified by # is correlated with the Chattahoochee River. Clusters represented by + and by blanks represent major tree types, while an 0 represents grasslands or fields. Most of

the other clusters may be correlated with manmade objects ranging from highway systems to residential areas and shopping centers. Clustering in this area represents for the most part information on the extent of man's activity. Table 1 lists the cluster means and the total ground area occupied by each cluster. Note that in Figure 11 the cluster + has a northeasterly elongation. This cluster will be discussed further in a later section.

Stone Mountain Test Site

Figure 12 shows the results of clustering in the Stone Mountain area. The cluster associated with the lake around Stone Mountain is depicted by the overprinted characters A, O, X, and V. Blanks and the characters 0 and = represent vegetation. Several clusters appear in areas where the Stone Mountain granite outcrops. Since each resolution element as used in this test represents approximately one half a hectare, the several clusters may represent different mixtures of trees and bare rock. These clusters are given by the symbols θ and ≠ . The symbol / seems to indicate quarries around Stone Mountain. The symbols , and # represent manmade objects. To the west of Stone Mountain one can identify patterns depicting the city of Stone Mountain represented mostly by the = cluster. Since the Stone Mountain area is extensively rural, and since there is little variation in the spectral signatures of manmade objects where they do occur in this area, more vegetation clusters were distinguished than in the Cumberland Mall area. Comparing the clustering map (Figure 12) with a geologic map (Figure 13) of the area, (Hermann, 1954), one finds little correlation between the two, apparently

indicating that this mode of analysis is not able to distinguish between the rock types in this area (Stone Mountain granite and Lithonia gneiss). However, these rock types have almost identical chemical compositions and weather to the same soil type. Most of the area is covered by vegetation growing on the same soil type, and as might be expected, no difference between the spectral signatures from the two rock types is detectable. Table 2 gives the cluster means and the area covered by the different clusters.

Macon, Georgia Test Site

Results of clustering in the Macon, Georgia area are shown in Figure 14. Ground truth information in the form of a landuse map of Macon is shown in Figure 15. When these figures were compared, several clusters were identified. The character ▣, which is an overprinting of O, A, X, and V, appears as a dark spot on the map and represents a cluster identifiable with the Ocmulgee River which winds through the area. Blanks and the character + are clusters that correlate with undifferentiated vegetation types, while the clusters /, ⊗, ⊖, and ≠ correlate with manmade objects in the business district. The cluster = seems to coincide with lower density urban areas such as single family homes. Table 3 shows the cluster means and the area covered by each cluster.

Carters, Georgia Test Site

For the Carters test site (Figure 16), clustering allows discrimination between vegetation types and differentiation of various manmade structures. This detectable difference in vegetation in this

area allows the indirect identification of the Cartersville fault. To the west of the fault much of the land is open and grassy. To the east of the fault the prevalence of wooded areas provides a distinct contrast which allows detection of the fault trace. Other vegetative differences east of the fault are also noted. These are in the form of patches elongated in a direction corresponding to both the topographic trends and one of the prominent geologic joint directions of the area (N 55 E) as mapped by the author (Figure 17). To the west of the fault, fields along the Coosawattee River are also enhanced. Clustering defines the watercourses of the area very well, including the creeks emptying into the Coosawattee. Abrupt differences are not found in tree types on opposite sides of the fault trace. It is the vegetative patterns rather than the spectral signatures which allow differentiation between late Precambrian and early Paleozoic metamorphic rocks east of the fault and the Paleozoic sedimentary rocks west of the fault in this example. The cluster means are shown in Table 4.

CHAPTER VIII

GEOLOGICAL RESULTS

So far in this thesis, we have been trying to establish the spectral characteristics of the spectral signatures and spatial distribution patterns of vegetation types which would give us some insight into the geologic structure of a region. For a few test cases the signatures were established by clustering and the results subsequently used to define classes needed for supervised classification as described in Chapter V of this thesis. The resulting classes were used in the application of supervised classification to other ERTS data sets, and an attempt was made to extract geologic information about the area covered. The outcome of this attempt is given below, where the results of analysis of the Douglasville area are considered.

Douglasville Area

Clusters developed in the analysis of the Cumberland Mall area were used as a basis for supervised classification in the Douglas County area west of Atlanta. This area has been under study by the Engineering Experiment Station of Georgia Tech as a test case for a landuse classification by use of ERTS data. The individual spectral means for each class are given in Table 1.

Up to this point in the study, no specific geological classes had been associated with the several different vegetative types observed in the Cumberland Mall area. However, the analyzed data from around

Douglasville, Georgia showed an elongation that corresponded to a spectral signature (indicated by the character + in Figures 10 and 18) of one of the tree types of the Cumberland Mall area. This particular vegetative type parallels the Chattahoochee River as it flows southwest from Atlanta. The trend is especially concentrated near the river and dies out both north and south of the river (Figure 18). In an effort to determine the cause of the elongation, soils maps (Figure 19) of the area were compared with the classified ERTS data. The comparison revealed that this vegetative class corresponds closely to the Louisa fine sandy loam soil. According to a soil survey of Douglas County (Wells et al., 1961), the Louisa series are shallow, strongly acidic soils that are somewhat excessively drained. The series is formed by the weathering of a mica schist. The organic matter in these soils is low, and they will support only loblolly pine, shortleaf pine, and underbrush. More than 95% of the Louisa series soils are covered by trees in this area. By virtue of this information the tree group extending from the Cumberland Mall area into Douglas county is inferred to be loblolly and shortleaf pine. Thus, since the soil group itself was not resolved but rather the specific vegetation that it supports, and since this vegetation seems to be indicative of this soil type, and indirect recognition of soil type has been achieved using ERTS data.

Figure 20 shows a geologic map (Crawford and Medlin, 1974) of the Brevard Shear Zone west of Atlanta. Here, the Chattahoochee River parallels the Brevard Zone, with the Brevard Zone lying north of the river. Comparison of Figures 18, 19, and 20 shows that the ERTS vegetation class loblolly and shortleaf pine (+ in Figure 18) and the

Louisa soil group both correlate well with the Garnet-Mica schist belt just north of the Long Island fault and the Chattahoochee River. The correlation of ERTS data to a vegetative type supported by the Louisa soil series that was derived from a mica schist implies that a major lithologic unit associated with a shear zone has been indirectly delineated by the automatic processing of ERTS data.

A second example of the usefulness of digitally processed ERTS data is the apparent correlation of several clusters of MSS data generated from an area near Lake Chatuge in north Georgia to exposures of ultrabasic rock. A geologic map of the area (Figure 21) was taken from a paper discussing the ultramafic rocks. Particularly the dunite, are characterized by thin or non-existent vegetative cover. A clustering analysis (Figure 22) was made of the ERTS data for the area to the east of the lake. The spatial location of several clusters was found to correlate well with the location of the ultramafic rocks. The limited ground truth information available indicates that the pattern detected is probably a lack of trees in a region that is othewise primarily forest. However, use of clustering seems to have again resulted in the indirect recognition of a particular rock type in Georgia.

Most of the spectral variation which differentiated the ultramafic rocks from other types in north Georgia occurred in channel 7 of the ERTS data. In this channel the rock clusters have very high reflectance and water has a very low reflectance. Vegetation falls in between these two extremes. Because of the unique nature of these two signatures in channel 7, a linear classification on only channel 7 of ERTS information was used. A one channel linear supervised

classification scheme was developed that would classify a resolution element into class 1 (water) if the spectral value in channel 7 was less than a value - S1 and into class 2 (rock) if the value was greater than S2 where S2 was greater than S1. The values of S1 and S2 were estimated from the previously computed mean and standard deviation of water and rock spectral signatures in channel 7. If a resolution element had a value between S1 and S2, it was assigned to class 3 (other) which contained primarily vegetation. This simplified linear classification scheme is 25 times as fast as either clustering or supervised classification which uses all four ERTS channels; however, it does not provide all the information that the latter analyses would provide. The main use for such a simplified scheme would be to quickly process MSS data covering a large area. Figure 23 shows an unscaled map of the Chatuge area generated by the linear scheme. Once a class had been tentatively established as related to the predominant rock in an area, further computer analysis or actual field examination could be used to verify or discount the classification. Since there are many lakes in north Georgia, the water classification could be used for the location of reference points on the computer map. The location of these classes could lead geologists to the general area of an exposure without resorting to preliminary fieldwork.

CHAPTER IX

ANALYSIS OF THE FAST FOURIER TRANSFORM OF BLOCKS OF ERTS DATA

If we have a two dimensional function g(x, y), then the Fourier Transform of the function is given by

$$G(K_x, K_y) = \int_{-\infty}^{\infty}\int g(x, y) \exp[-2\pi i (K_x x + K_y y)]\, dxdy$$

where K_x and K_y are wave numbers of waves normal to the x and y directions respectively.

The Fourier transform has been extensively used in many fields of physics and engineering. It converts distance versus amplitude data in the space domain to wavelength (or its inverse - wavenumber) versus amplitude data in the wavenumber domain. The wavenumber domain contains the same information that exists in the space domain but the information is expressed in terms of the amplitude and phase of component sinusoidal waves.

In contrast to the continuous form of the above equation, ERTS data are finite and discrete. In each channel a single scan of the ERTS sensors produces 3240 resolution elements, each of which has a single integer value associated with it. Hence, ERTS data are not continuous, and the Fourier integral transform, as such, cannot be computed. The effects of the finiteness (i.e. truncation) and discrete sampling (i.e.) digitization) will be illustrated below.

The Fourier integral transform of a continuous pure cosine wave is an impulse in both the plus and the minus directions at the frequency of the wave on the frequency axis (Figure 24). Assuming a finite length (X) of data limits the data in the space domain and causes the spectral data to become digital (i.e., discrete) at wavenumbers of $n\Delta k = 1/X$. The valid wavenumbers are $n\Delta k$, where n is an integer; and provided that the space domain is continuous, the value of n can become infinite. By constraining the data to a finite interval, assumptions are made concerning the continuity of the function at the edges of the interval. Ideally, the function should be contained entirely within the sample interval; however, this is not always possible. Often continuity is approximated by smoothing the edges or removing a residual. In effect, when the data are truncated they are assumed to be cyclic such that the last point at $x \sim X$ is followed by the first point at $X \sim 0$. This cyclic character often will introduce artificial high-frequency information if the ends do not meet. For example, if we consider a ramp function in the space domain, (Figure 25) the transform of this function distorts what should be a long wavelength trend into one with large high frequency components. The truncation of this trend results in a severe limitation on the interpretation of wavelengths longer than the sample interval. Because of the high frequency nature of ERTS data this effect is not particularly detrimental to the interpretation of ERTS data.

Discrete sampling of a continuous function also effects the interpretation of the data. Figure 26 shows a sampling of a continuous cosine function with a sample interval of $\Delta = 2\pi/20$. These data

adequately represent the nature of the cosine function. However, if the sampling interval was increased to $\Delta = (9/5)\pi$ (Figure 27) the sampled data give an erroneous description of the original function. For this case a cyclic function of frequency w is represented digitally by a function with a much longer wavelength. This effect is called alaising (Bergland, 1969) and it can limit the usefulness of the transform. ERTS data, nevertheless, are sampled at the highest frequency possible for the MSS (i.e., w~1/resolution element size), and the spectral data obtained represent the integral over the area of the resolution element. For a discrete spatial data set with n samples the highest frequency possible is n/2 cycles per unit distance. This would be the case if the function were sampled only at each maximum and minimum of the function. The highest wavenumber, 2/n, is called the Alding wavenumber and corresponds to the folding frequency of the transform.

A rapid method for the computation of a digital Fourier transform is discussed by Cooley and Tukey (1965). This method is called the fast Fourier transform (FFT). The FFT may be considered an approximation to the continuous Fourier transform (CFT) or the exact representation of a finite dimensional vector Fourier series transform.

Almost all techniques for digitally processing ERTS data utilize only spectral radiation intensities and ignore information about the classification of resolution elements surrounding a candidate resolution element. The transform technique should allow us to use some of the information about the spatial relationship of the resolution elements in an ERTS scene. One specific use of the FFT that this study

investigates is its capability for linear trend analysis. Several sample data sets are considered to illustrate the FFT lineation detection capabilities.

The first sample data set consists of a function of the form $z = A + \cos(2\pi fx)$ with $f = 2/x$ which represents a corrugated surface with ridges parallel to the y axis (figure 28). A is an offset along the z axis of 1/2. If this function is sampled at a rate of one sample for each integer increment in the x and y directions for 32 elements, a 32 x 32 element digital representation of the function may be obtained as is shown in Figure 29. This data set is sampled at exactly 1/2 the wavelength of the ridges in the x direction. In the y direction the wavenumber of the ridges is zero. The 'star' diagram of the magnitude of the FFT for this data set is shown in Figure 30. Values to the right of the origin in the positive K_x direction are magnitudes of wavenumbers corresponding to wavelengths of waves which are normal to the x direction in the space domain. At the far left is the Alding wavenumber. Along the positive K_x axis the wavenumbers increase from dc to 1/2 cycle per unit distance. The K_y axis represents wavenumbers in the y direction of the original data set. Since the only periodic variation in this data set is in the x direction at a wavenumber of 1/2, the FFT star diagram shows large values for the dc component of the FFT and for components at a wavenumber of -1/2 cycles per cell in the x direction.

A noncyclic linear pattern which strikes parallel to the y direction can be approximated by a sum of wavenumbers in the K_x direction. Therefore, in the wavenumber space a linear function parallel to

the y axis is represented by a linear pattern of values in the K_x direction. In general, the transform of a noncyclic linear pattern is a linear pattern orthogonal to the original linear noncyclic trend. A sample noncyclic linear trend is shown in figure 31. Figure 32 is the FFT of this data set. The data values in the upper righthand corner and lower lefthand corner are due to truncation effects.

As an example of the use of the FFT algorithm on a ERTS data set, the FFT was applied to a subset of data in the Cumberland Mall area of Atlanta. Noticeable linear patterns in this set are I-75 and ridges paralleling the Brevard Shear Zone (see Figure 37). The ridges trend predominantly northeast and are not readily recognized in clustered data. Figure 33 shows an FFT star diagram of channel 7 for this area. As discussed above, a linear pattern through the center of the star diagram is orthogonal to a noncyclic linear pattern in the original data. In this figure there is a distinct trend which strikes northwest; this indicates a linear pattern in the ERTS data of northeast which corresponds to the Brevard zone. Figure 34 gives the FFT of channel 4 of the same area. A definite trend seems to head northeast signifying a linear trend in the LANDSAT data to the northwest corresponding to I-75.

Thus, in two artificial data sets and one ERTS data set the capability of the FFT alorithm to enhance linear trends in demonstrated.

Figure 35 is an FFT (channel 7) of a part of the mountainous area to the east of the Cartersville fault. Possible trend directions include north-south, NNE, and NE. These directions correspond well to joint patterns (Figure 36) in the Carters area. The north south

trend is due to the scan lines in a ERTS scene.

CHAPTER X

DEVELOPMENT OF A SPATIAL CLASSIFIER

As emphasized in the preceding section, a prime drawback of most classification methods for earth resources data is the limitation of the analysis to point-by-point classification. Each classification of a resolution element is computed as if it were independent of any information about the surrounding area. Because of this limitation much information is neglected that might improve classification results. By considering only one point at a time, a linear trend in the data could not be detected automatically or considered in the numerical analysis. A pattern detection algorithm for linear trends using the FFT was discussed above. This same algorithm can be made part of a general analysis scheme which incorporates spatial information into the classification process. One such scheme is detailed below.

Fof FFT analysis, a 32 by 32 square of resolution element in one band of the spectrum is transformed into the wave number domain. The same information in the wave number domain is a symmetric 32 by 32 element matrix which can be ordered so that the center element is the dc (0 wave number) component and the wave numbers increase in the K_x and K_y direction away from the center. This star diagram was illustrated above. Information about spatial trends in the area of the input image can be inferred from the star diagram of the FFT. Gross properties of the transform may be characterized by computing numbers from an integral

of the wave number magnitudes in different regions of the wave number space. For the study of linear trends four regions could be chosen based on 45 degree segments of the righthand side of the star diagram. Integrals in the wave number space excluding the near dc components could be calculated for each segment (Figure 38).

Implementation of this scheme requires that a new data base be constructed from the original data base of earth resources data. These data may be ERTS or aircraft MSS digital data. As each resolution element is considered, a block of data from one ERTS channel centered on that element is transformed by the FFT. The four integrals described above are then returned and stored as the lower four elements in a five dimensional vector for each component of the original ERTS vector with the initial component being the original spectral vlaue in one data channel for the resolution element. Thus, four components of spatial information related to the magnitude of linear trends in four directions would be stored for each spectral value. For each resolution element considered, instead of four channels of data to be used in a classification, now twenty channels of data would be available. A channel selection algorithm or a transformation such as principle axis (Duda and Hart, 1973) is then used to determine which channels will be used in the classification.

Once the new data base is constructed, the data could be processed by clustering to yield computer maps of the area. These maps should be more descriptive than those obtained by the simple spectral classification since spatial information would be included in the

classification.

The integrals in the wavenumber domain illustrated above are designed to distinguish directional preferences in the data. In this study another integral in the wavenumber domain was designed to distinguish whether the signals were coherent, i.e.--smooth with long wavelengths, or predominantly rough with short wavelengths, or in between these two extremes. Figure 39 shows the division of the star diagram into three sectors which measure the magnitude of long, intermediate, and short wavelengths.

Figure 40 shows two lines of ERTS data across the Atlanta area that are used as preliminary test sites for the spatial classification technique illustrated in Figure 39. FFT's of a 32 by 32 element block centered around each ERTS element along the lines were calculated. Channels four, six, and seven of ERTS data were considered individually. Line one goes from a low density residential area in northwest Atlanta to the Buckhead and Lenox Square commercial centers. A distinct variation in the summed FFT channel four in Figures 41 and 42 is noted in the transition from low density housing to commercial areas by comparison to figure 43 which is a computer map of the Atlanta area generated by spectral classification.

The curve in Figure 41 is at the same scale as the computer map. Peaks in the left side of Figure 41 correlate directly with the I-285 freeway area and 1-75 freeway area. The large peaks to the right of Figure 41 represent the intensely commercial regions of Buckhead and Lenox Square. The size of the peaks seems to be in direct correlation with the amount of commercial activity. The minimum to the left of the

largest peak represents an area of low density housing.

Figure 42 shows data over the same line as shown above, but displayed by 45 degree segments of the righthand side of the FFT as discussed above. This plot has two interesting anomalies that were not noted in the discussion of Figure 41. Segment 3 has a large peak at the left side of Figure 42. Since lineations in the FFT are orthogonal to lineations in the original picture, this peak would represent a linear trend in the original picture function of North-Northeast with respect to the data (I-285). The second anomaly is the peak in segment 1. This could correlate to a linear trend in segment 3. By comparison to Figure 43, it can be seen that this peak correlates well with I-75 which tends West-Northwest.

Figure 44 is a plot of summed FFT magnitude in the low, medium, and high frequency ranges of channel 6 of the ERTS data over line 2. This line passes from highly commercial regions through a warehousing region to another highly commercial region. The peak in the high frequency region on the left side of Figure 44 is due to a north trending freeway and commercial area along I-285. The peaks in the center of the figure represent concentrated commercial areas.

Since a FFT has to be computed centered on each resolution element, one drawback of this method is the long computation times needed to derive the new spatial data base. For ordinary digital analysis of large data sets this expense would certainly be prohibitive; however, analog processing techniques offer an alternative. A classification scheme such as described above could be combined with an optical Fourier transform of an earth resources data set. Since the Fourier transform

is a natural operation in optics, a coherent Fourier transform processor such as described by Shackelford and Walsh (1975) could be used to take transform information directly as the laser scans an aircraft or spacecraft photograph. Using this technique a digital data base with spatial information can be developed. This digital data can then be used in the previously described classification systems. A two element disk detector, one half of which consists of radial optical detectors and the other of concentric detectors is used in the system.

Discussing a similar optical system, Gramenopoulos (1974) studied several large topographic units. For large units Gramenopoulos suggested using optical Fourier transform data as an addition to ERTS spectral data for general terrain classification. Ulaby and McNaughton (1975) also suggested the use of an optical Fourier transform for detecting linear features.

CHAPTER XI

CONCLUSIONS

Several mathematical techniques have been applied toward the analysis of ERTS data for geologic information. Useful information can be gained by these types of analysis in the definition of geologic structure as well as in other disciplines. For example, in the clustering analysis applied to the Georgia test sites, clusters were generated which corresponded to different types of manmade construction which include highways and industrial complexes. Also, since open water can be easily detected, the information should be useful for location of watercourses and lakes for water resource studies. The manmade objects clusters can be used in land use studies, and the clusters locating watercourses can be used in the development of a data base for a watershed management program. Rock exposures such as Stone Mountain and quarries can be identified. Vegetation differences can also be detected which allow separation of grasslands and different tree types. In marsh areas, wetlands can be roughly deliniated by the lack of manmade activity in an area surrounded by housing and the presence of water throughout the wetland corridors.

The supervised classification technique for this study was used to extend the preliminary clustering results for a test area (Cumberland Mall) to other parts of the state. This method allowed discrimination of tree types in Douglas County, and led to the correlation of the

loblolly - shortleaf pine tree group in the classification results to the Louisa soil group. This soil group and the vegetation type were subsequently related to a mica schist geologic zone along the Brevard Shear Zone. This type of indirect information may be applicable in other parts of the state for locating geologic structure.

FFT analysis of ERTS data can be useful in detection of linear trends as shown by the recognition of manmade linear features and natural geologic linear trends. This technique is used to provide some spatial information to be used in conjunction with the ERTS spectral information in a unique spatial classification scheme developed in this thesis.

This study has proved, therefore, that the digital analysis of ERTS data for geologic information is feasible in the Southeast even though indirect means normally must be used to extract the information. The limitation of heavy vegetative cover in the Southeast makes these automatic techniques less applicable here than in a sparsely vegetated region; however, the use of indirect information has proved to be an acceptable method for geologic recognition.

CHAPTER XII

RECOMMENDATIONS FOR FURTHER ANALYSIS

The methods employed in this thesis have been useful in the determination of how appropriate digital processing of ERTS data is to the location of geologic structure in the Southeast. However, there is much work to be done in several areas that may enhance this capability. For example:

1) The spatial information classification algorithm developed in this thesis should be tested on more ERTS data.

2) Supervised classification should be employed along extensive sections of regional features such as the Brevard Zone to determine how universally applicable the techniques discussed in this thesis are.

3) Cost effective classifiers such as linear classifiers should be used for massive data reduction while the other methods should be restricted to specific regions.

TABLE 1

CLUSTERING MEANS - CUMBERLAND MALL

Cluster	Symbol	Size	Hectares
1	.	129	79.9
2	,	3479	2155.5
3	+	1606	955.1
4	=	564	349.4
5	≠	354	219.4
6	0	1420	885.8
7	⌀	264	163.6
8	⌀	93	57.5
9	/	125	77.3
10	▩	41	25.1
11	-	17	10.5
12	#	91	56.3
13	%	485	300.4
14	<	52	32.0
15	>	6	3.7

	1	2	3	4
1	.000	233.895	233.248	227.335
2	.000	241.690	241.033	232.134
3	.000	227.238	222.008	218.308
4	.000	239.284	235.375	235.323

	5	6	7	8
1	223.000	229.617	233.082	230.388
2	223.648	235.377	240.763	237.818
3	218.031	224.655	233.450	213.327
4	237.784	239.014	244.077	230.294

	9	10	11	12
1	219.191	216.478	207.507	233.631
2	215.445	206.338	203.916	242.345
3	213.646	206.408	204.897	241.238
4	236.967	234.525	233.170	249.463

	13	14	15
1	225.333	221.039	232.283
2	228.399	222.673	239.372
3	223.604	225.381	224.036
4	240.232	242.919	223.887

TABLE 2

CLUSTERING MEANS - STONE MOUNTAIN

Cluster	Symbol	Size	Hectares
1	.	123	76.1
2	,	1225	759.1
3	+	2236	1385.4
4	=	764	473.4
5	*	260	160.7
6	0	889	550.6
7	ø	55	34.0
8	Ø	5	3.1
9	/	16	9.9
10	R	244	151.0
11	-	106	65.6
12	#	237	146.8

	1	2	3	4
1	.000	230.023	233.145	226.833
2	.000	236.377	240.949	231.413
3	.000	224.652	227.993	222.332
4	.000	238.616	239.997	238.286

	5	6	7	8
1	221.142	233.082	211.258	193.186
2	223.949	240.752	209.737	187.286
3	219.298	223.658	209.898	191.613
4	238.016	236.575	235.013	227.887

TABLE 2 (Cont.)

	9	10	11	12
1	202.609	233.682	227.182	231.077
2	198.856	242.339	232.787	238.537
3	201.564	241.644	215.380	231.875
4	232.048	249.592	233.168	243.480

TABLE 3

CLUSTERING MEANS - MACON

Cluster	Symbol	Size	Hectares
1	.	126	78.1
2	,	4609	2855.8
3	+	1963	1216.2
4	=	884	547.8
5	▪	71	44.0
6	0	15	9.3
7	⊖	8	4.9
8	⊗	397	245.7
9	/	519	321.5
10	▩	144	89.1

	1	2	3	4
1	.024	229.269	225.555	221.273
2	.021	238.211	232.077	225.668
3	.034	226.686	223.799	220.396
4	.041	240.034	239.157	238.509

	5	6	7	8
1	215.500	210.610	196.556	216.616
2	214.267	205.272	187.427	220.145
3	212.225	203.195	187.880	220.230
4	236.041	232.515	225.795	239.990

TABLE 3 (Cont.)

	9	10
1	221.100	226.445
2	227.107	234.574
3	225.461	232.068
4	241.901	244.910

TABLE 4

CLUSTERING MEANS - CARTERS

Cluster	Symbol	Size	Hectares
1	.	125	77.3
2	,	4916	3046.0
3	+	2709	1678.5
4	=	380	235.2
5	✚	281	174.1
6	0	18	10.9
7	⊖	137	84.6
8	⊗	59	36.6
9	/	63	39.0
10	▣	48	29.7

	1	2	3	4
1	.042	233.954	234.434	234.815
2	.046	241.295	241.756	242.416
3	.028	221.726	226.490	232.987
4	.027	235.197	238.501	243.288

	5	6	7	8
1	231.249	230.429	229.012	224.026
2	238.561	237.711	234.265	227.062
3	217.917	212.101	223.936	227.542
4	233.240	229.415	238.972	244.025

TABLE 4 (Cont.)

	9	10
1	221.754	228.027
2	221.265	233.461
3	220.719	235.323
4	240.997	247.987

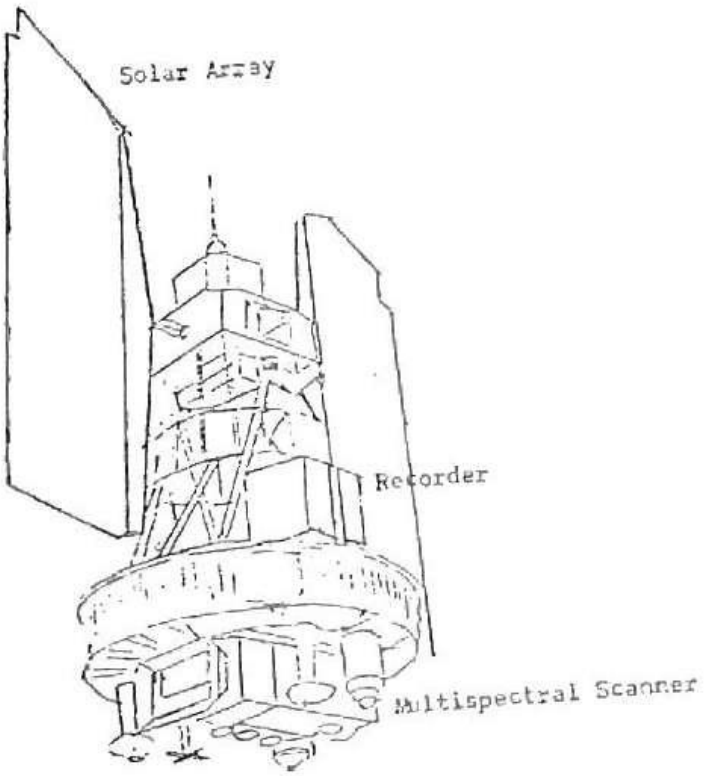

Figure 1. ERTS

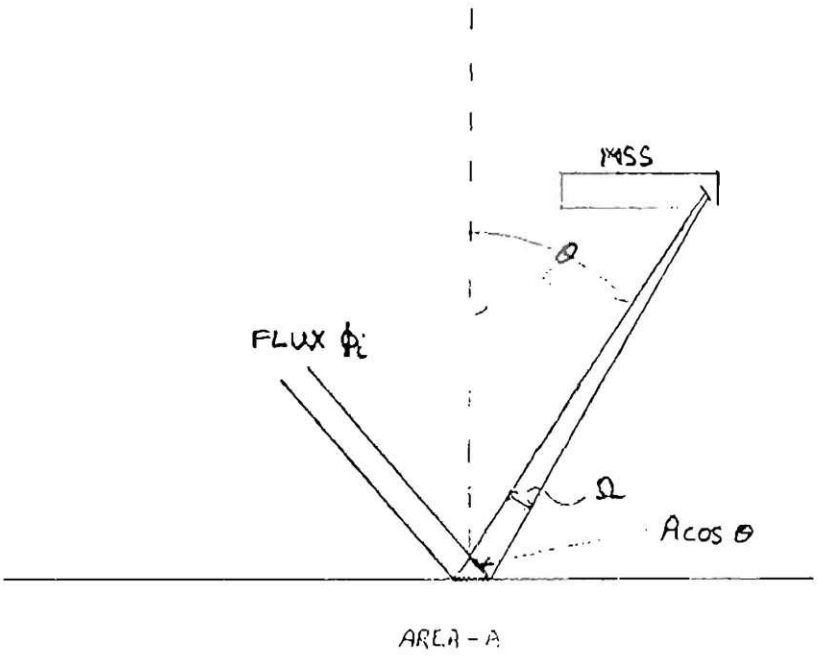

Figure 2. MSS Irradiance Configuration.

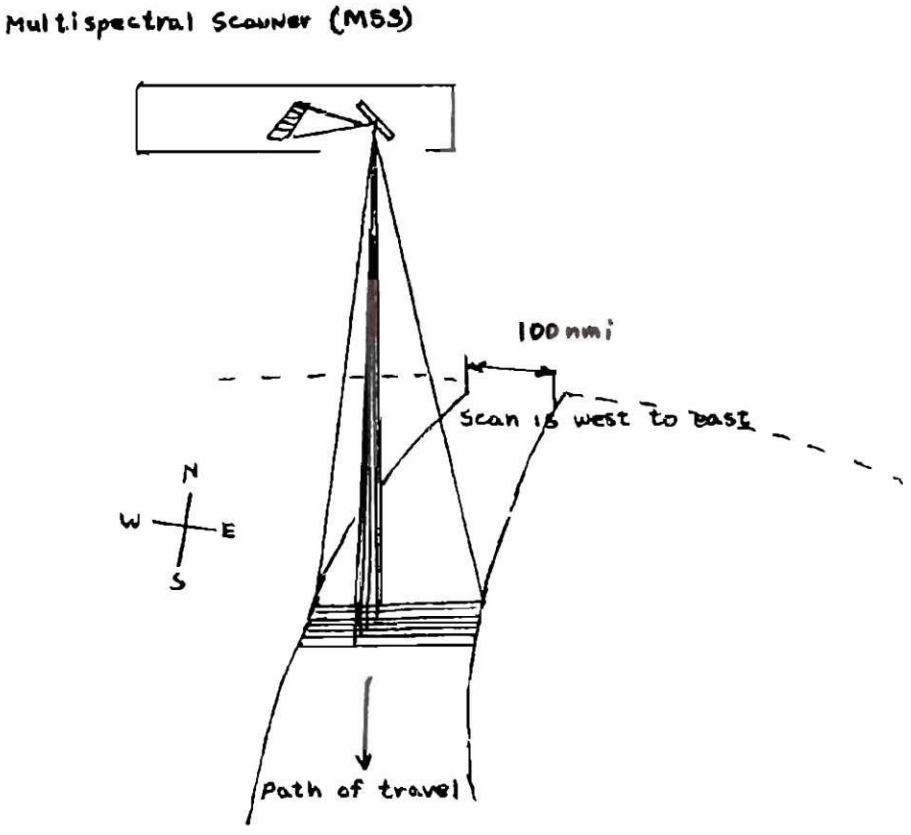

Figure 3. ERTS MSS Scan.

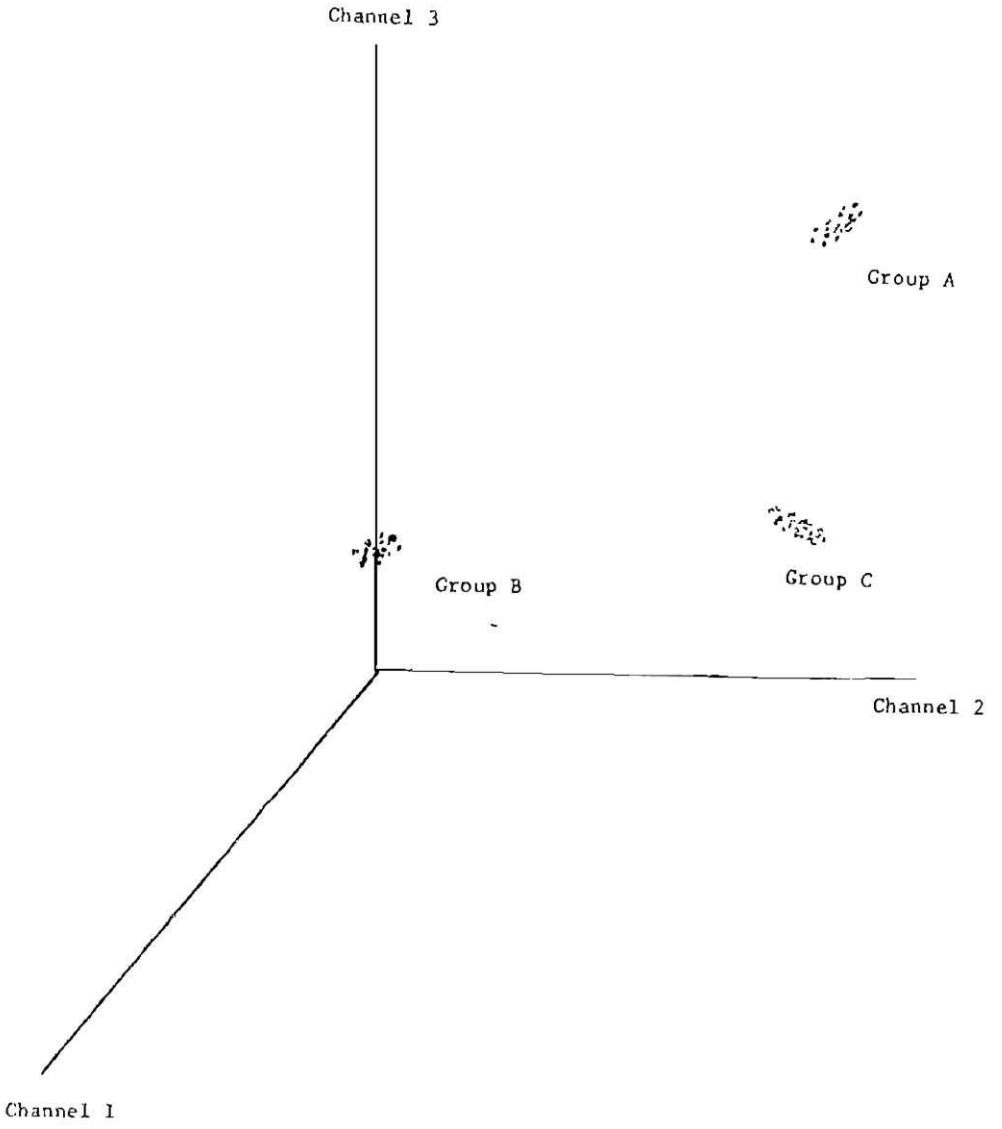

Figure 4. Location of Vector Group.

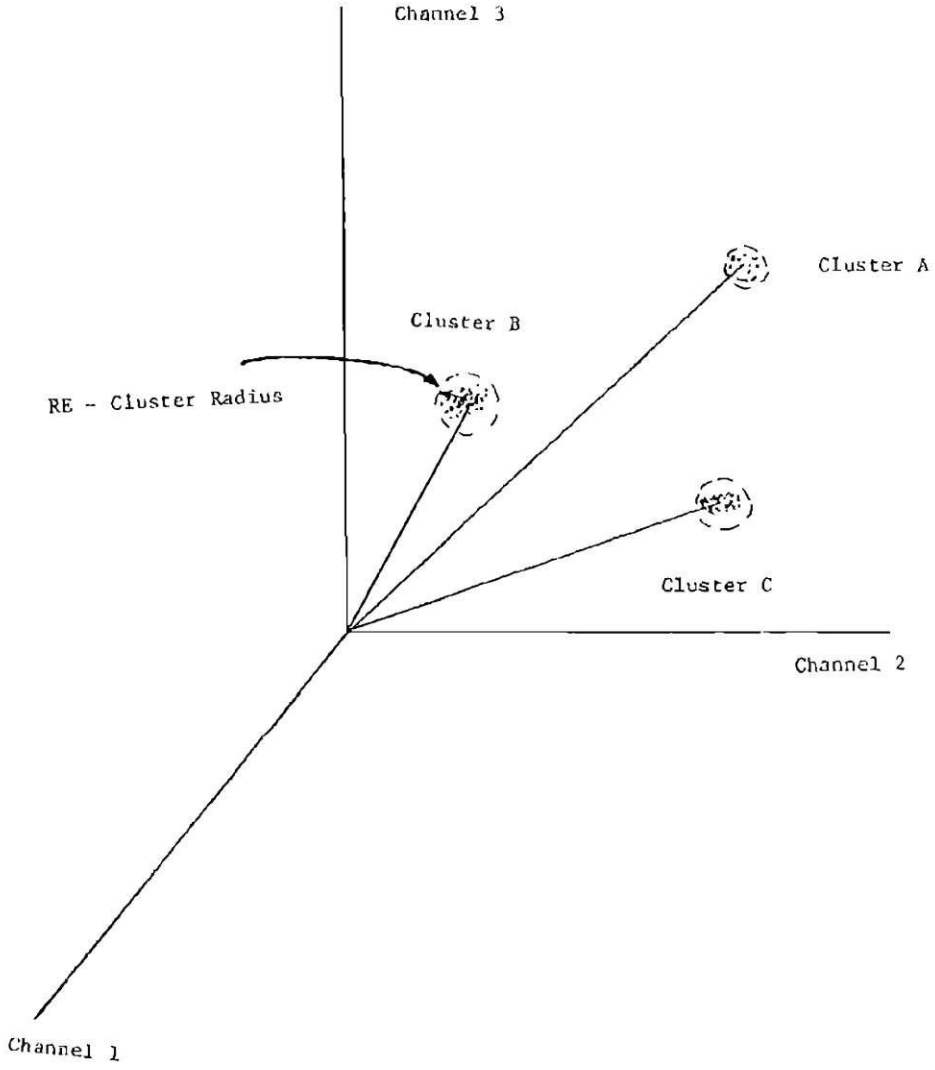

Figure 5. Cluster Location.

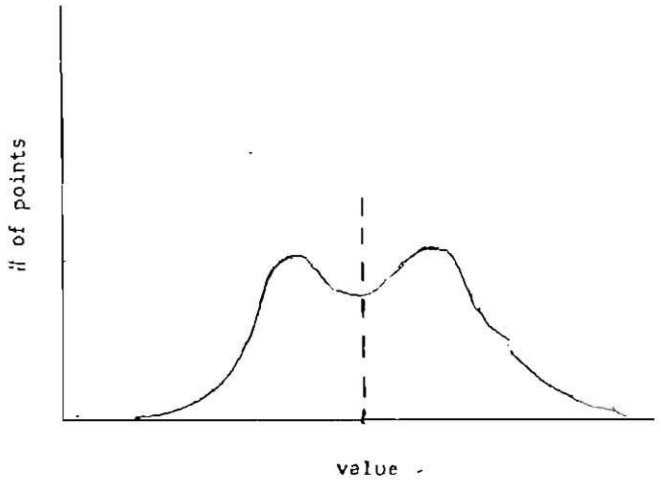

Figure 6. Bimodal Distribution.

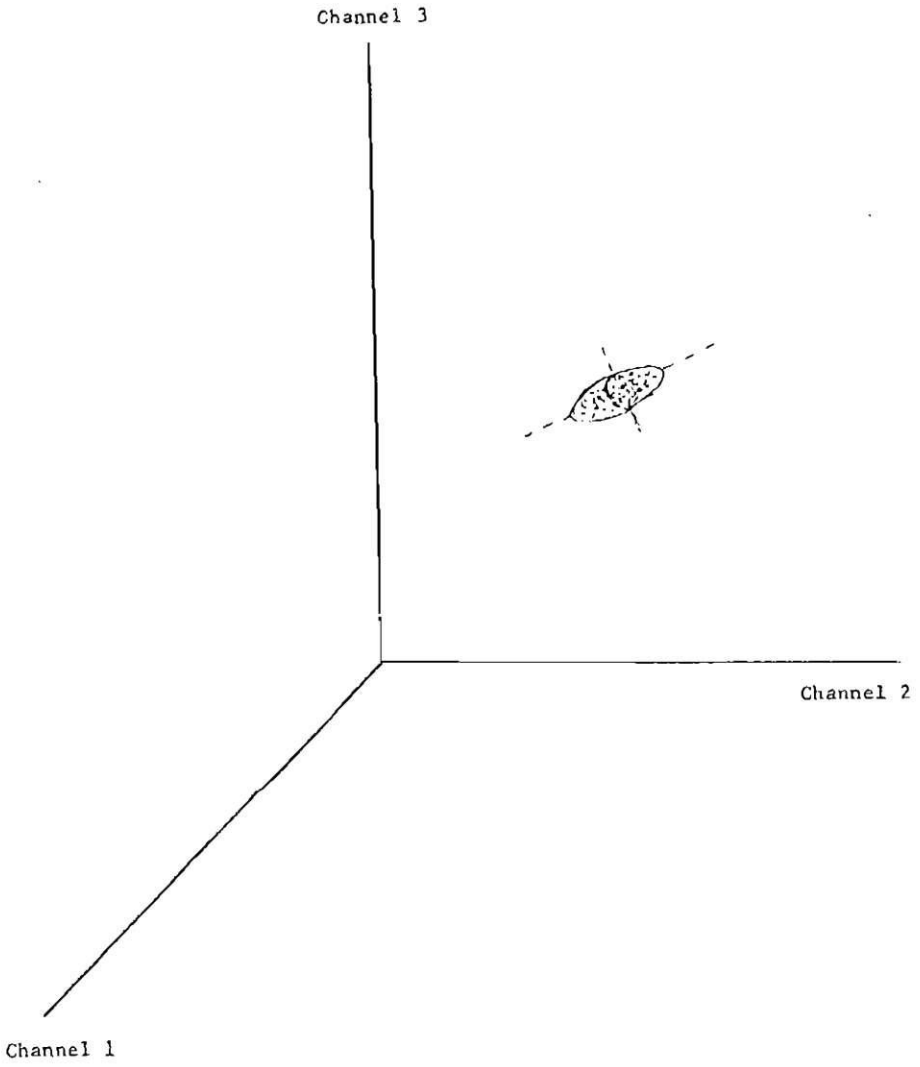

Figure 7. Three Dimensional Normal Distribution.

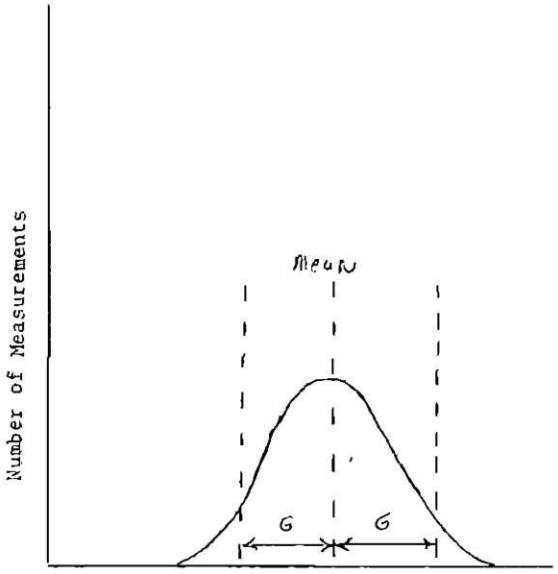

Figure 8. One Dimensional Normal Distribution.

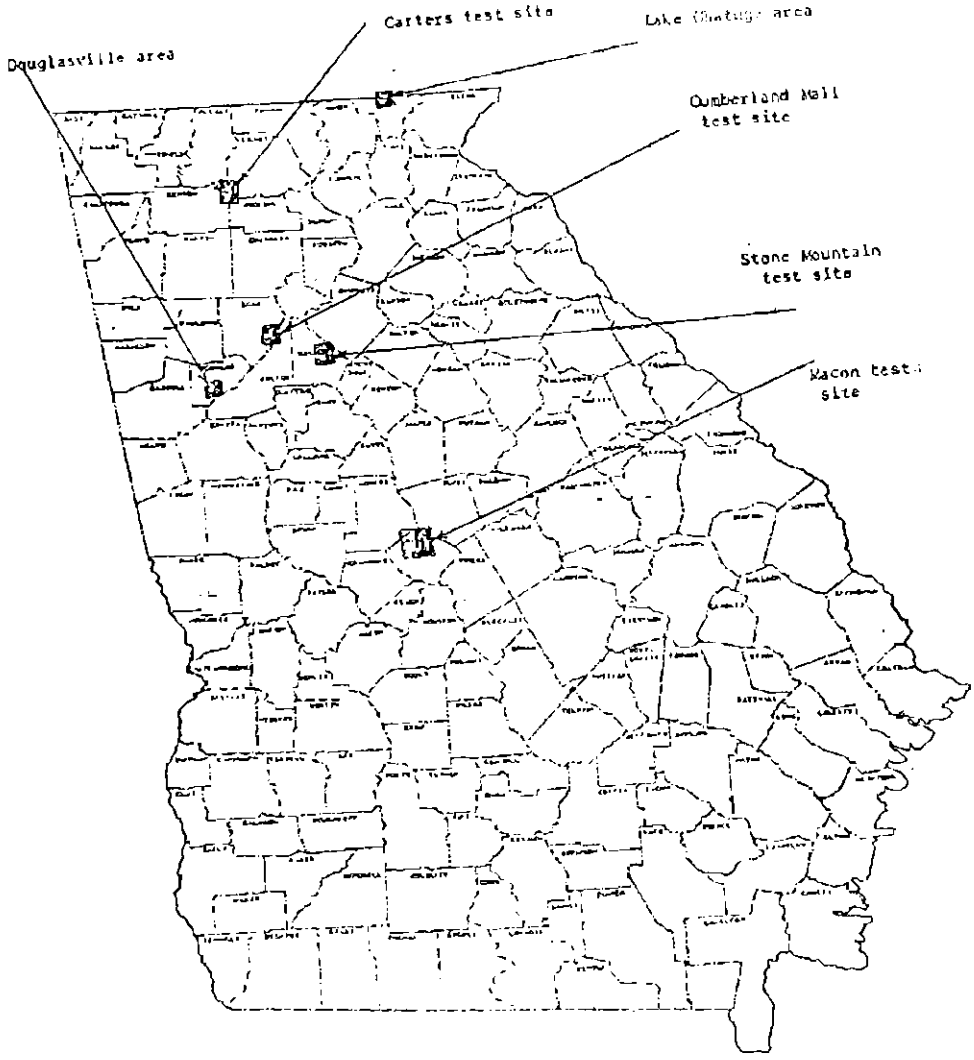

Figure 9. Location of Test Sites.

Figure 10. Clustering Data for the Cumberland Mall Area.
Boxed areas in Figures 10 and 11 Correspond to the Same Areas.

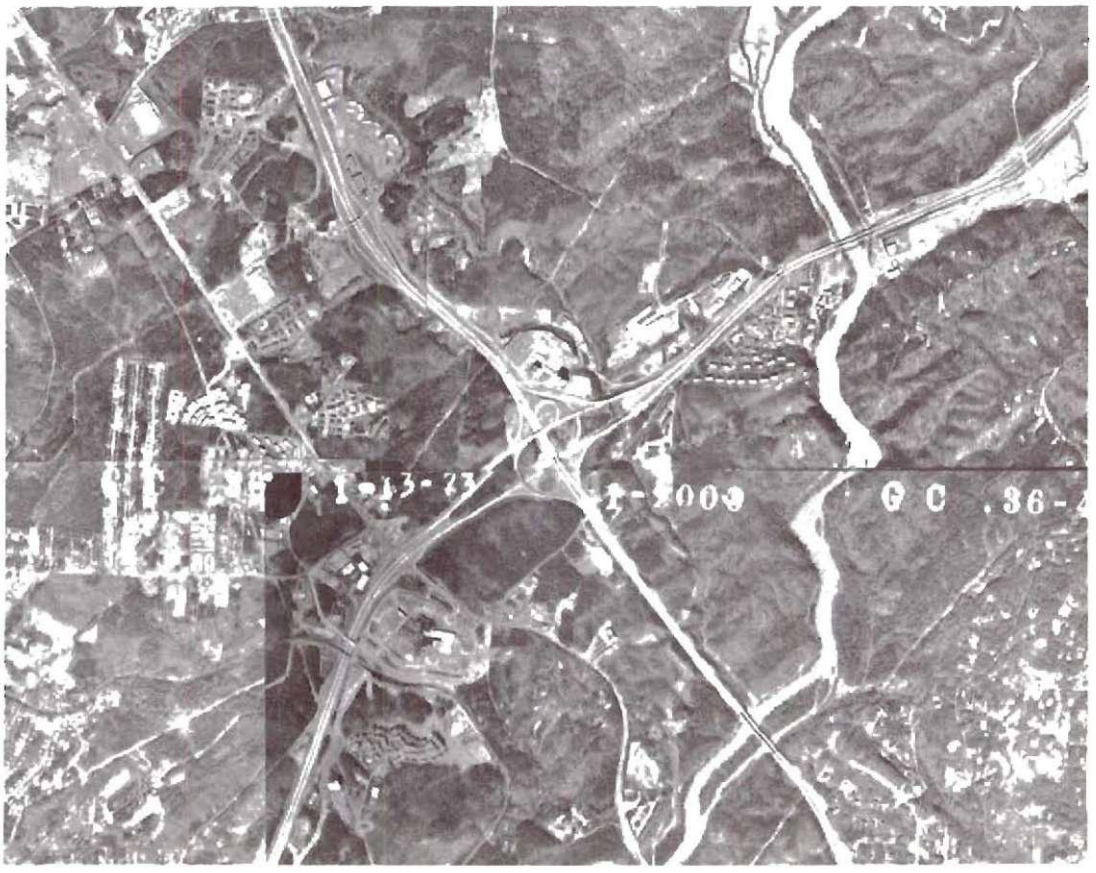

Figure 11. Aerial Photograph of the Cumberland Mall Area.

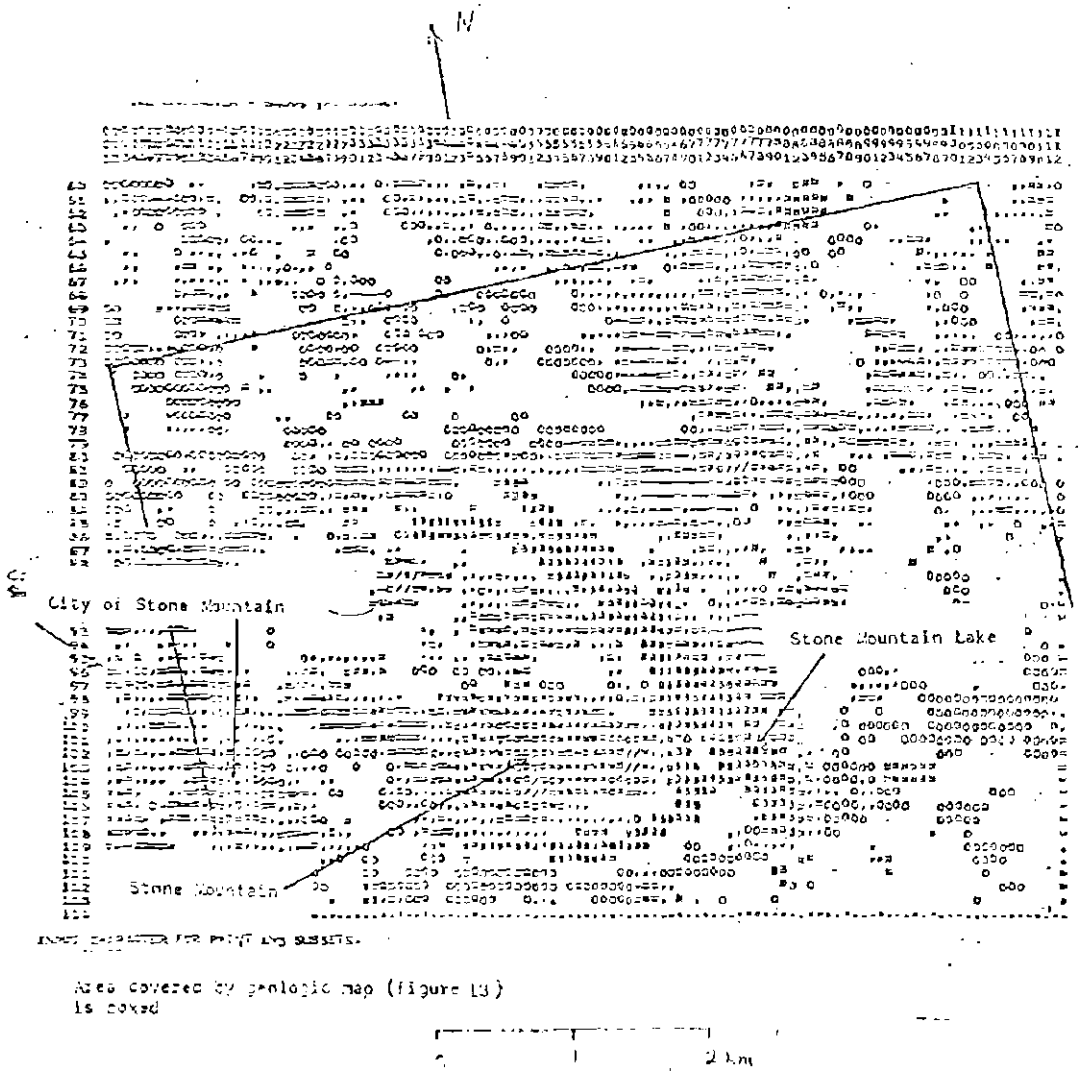

Figure 12. Clustering of Stone Mountain Area.
Covered by Geologic Map (Figure 13) is boxed.

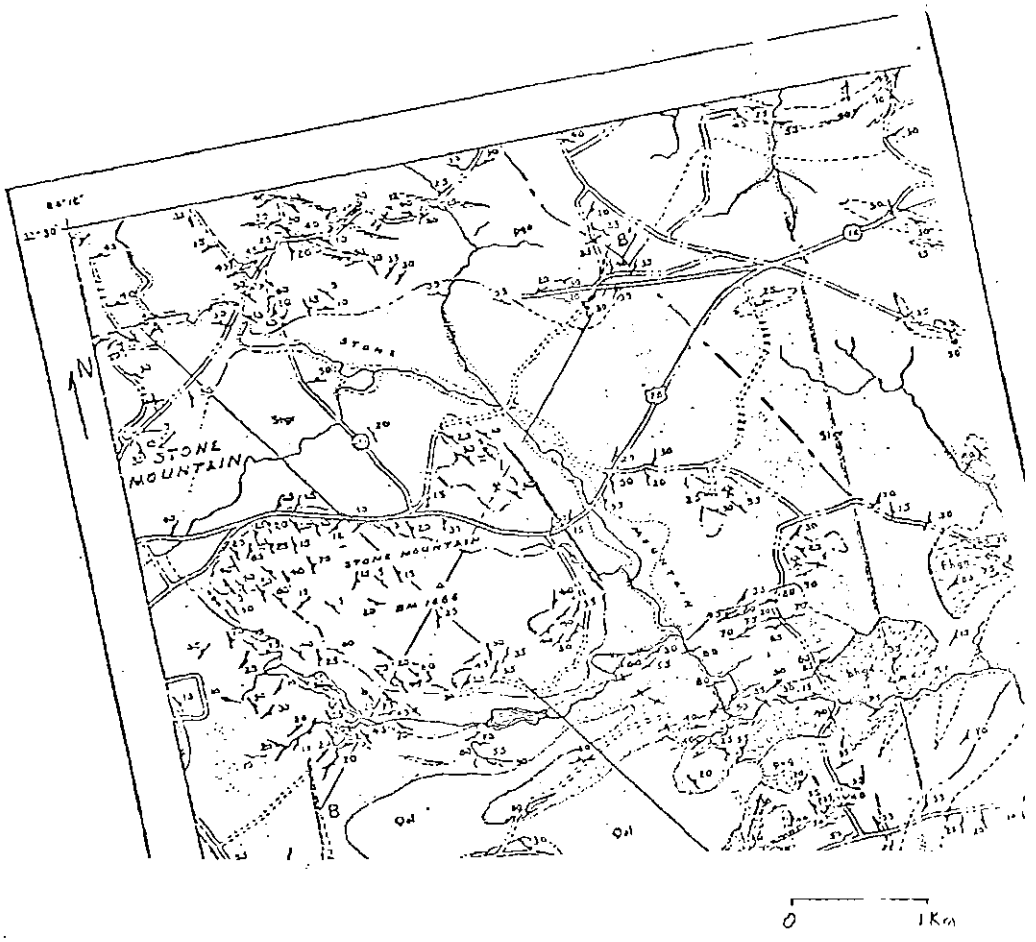

Figure 13. Geologic Map of Stone Mountain Area.

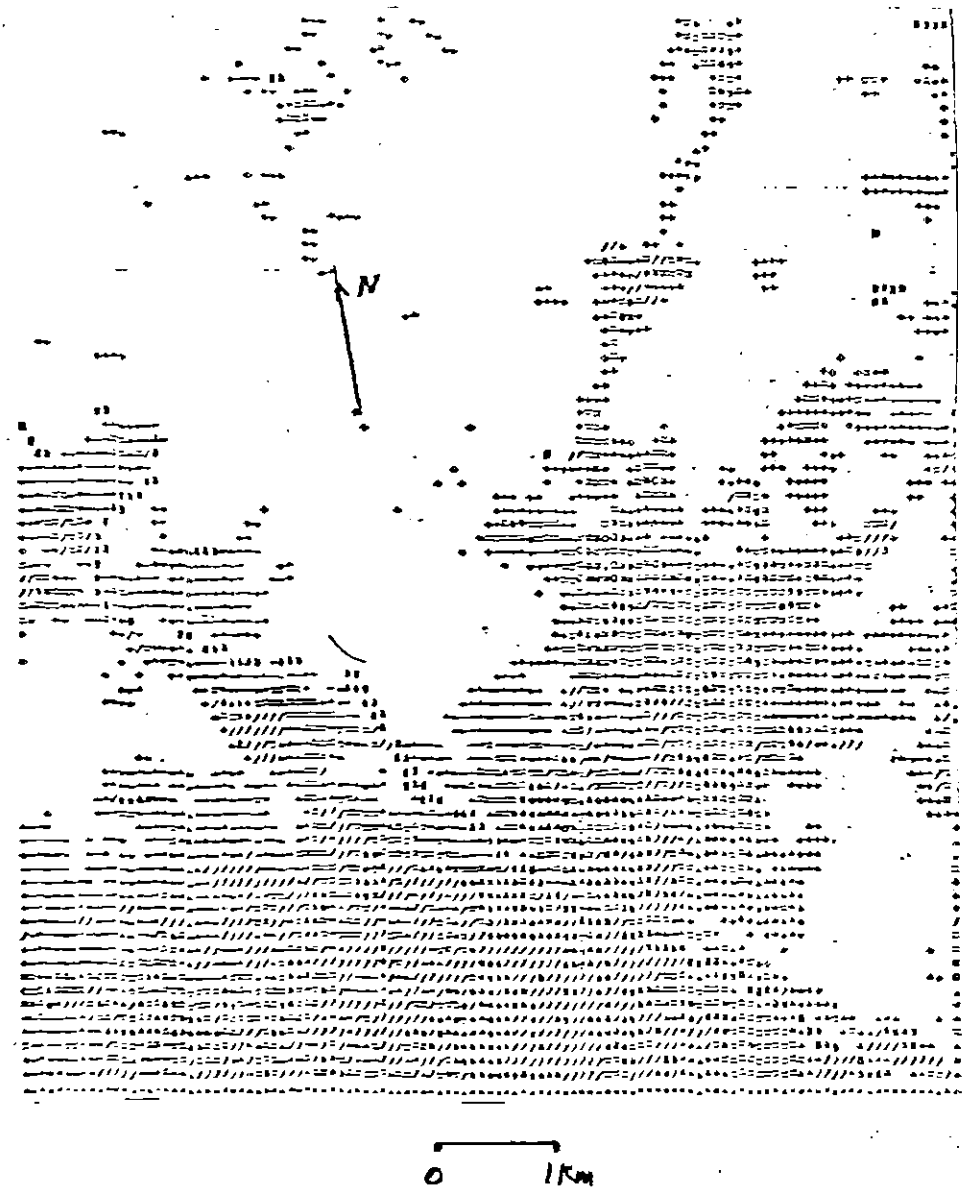

Figure 14. Clustering of Macon Area.

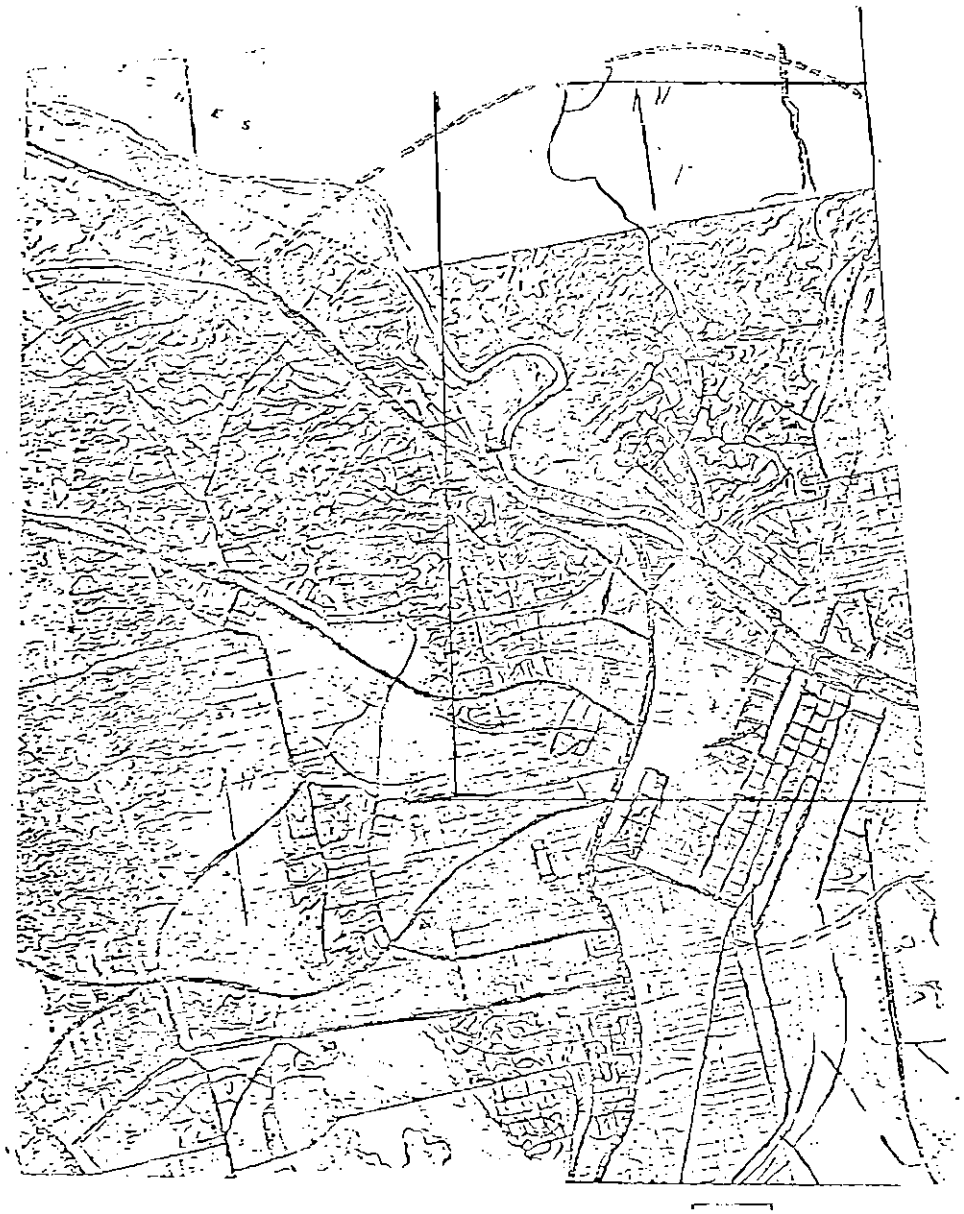

Figure 15. Landuse Map of Macon. Boxed Area is Clustered Area in Figure 14.

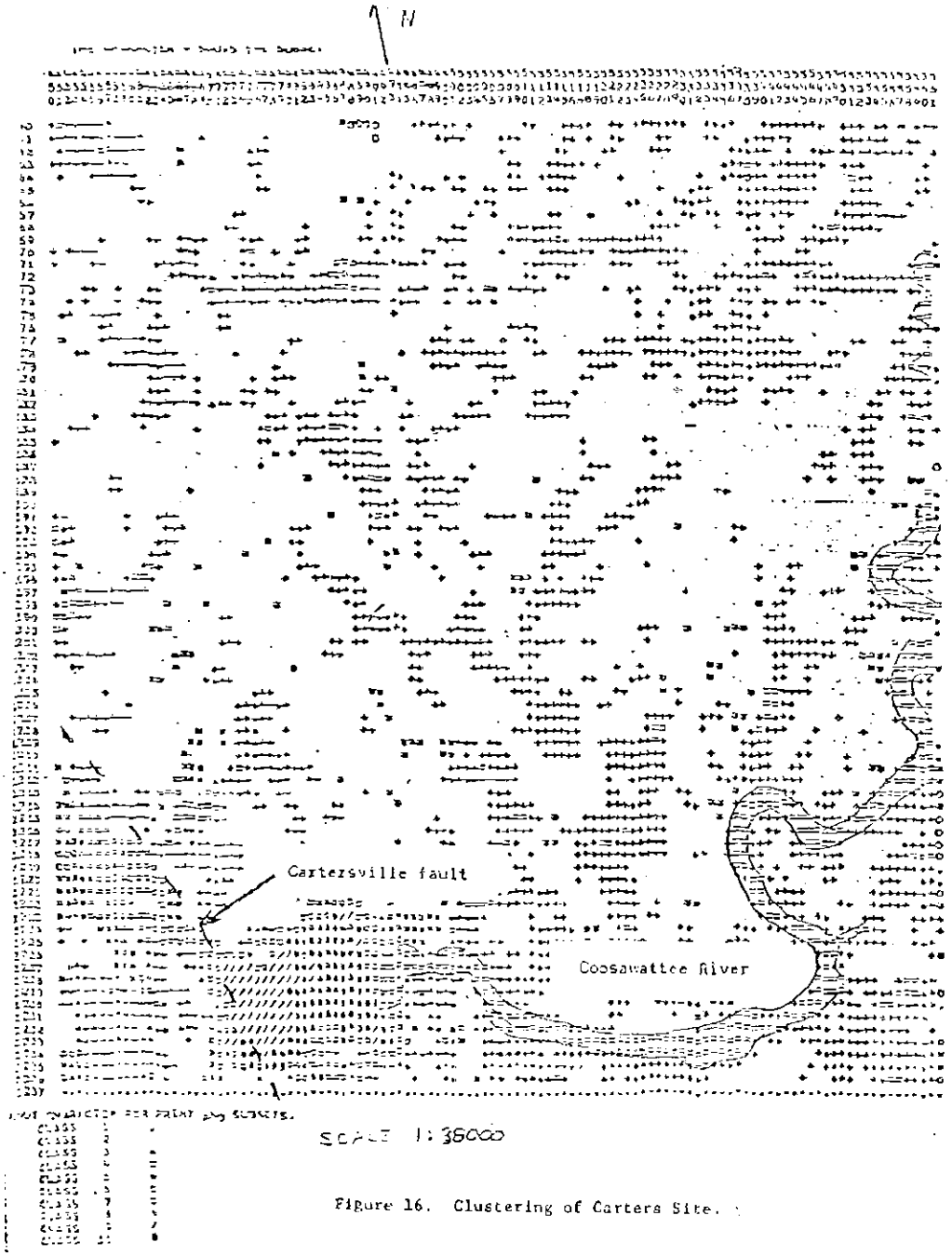

Figure 16. Clustering of Carters Site.

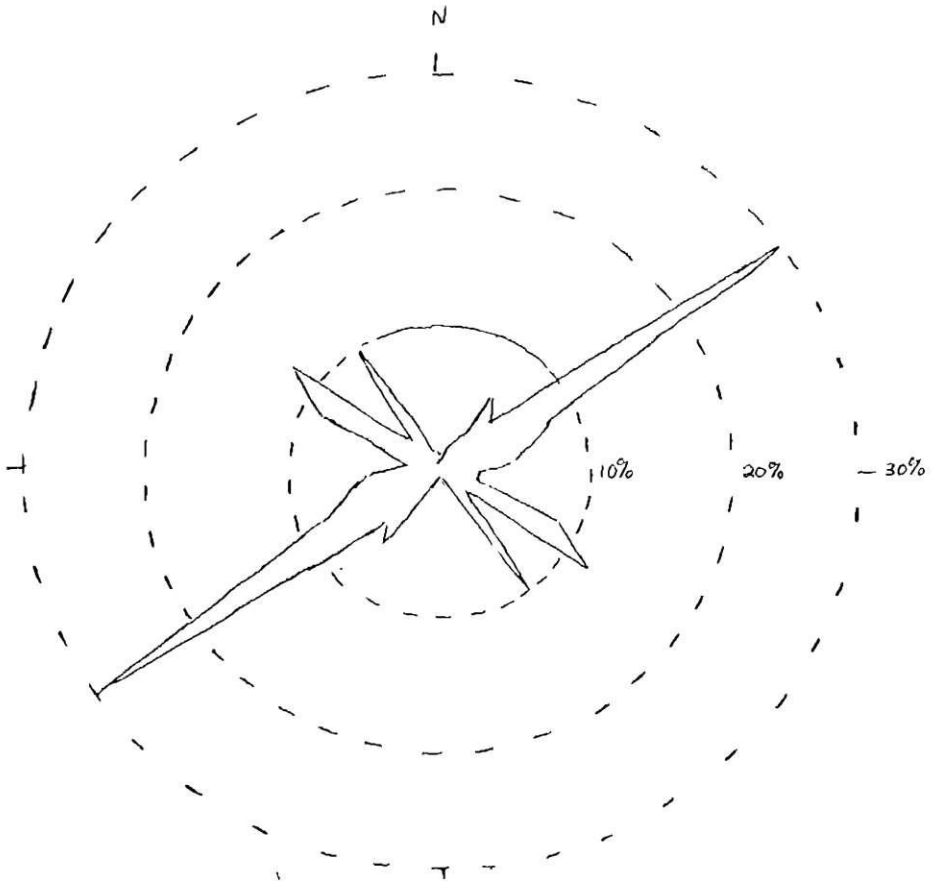

Figure 17. Joint Rose Diagram - Carters, Georgia.

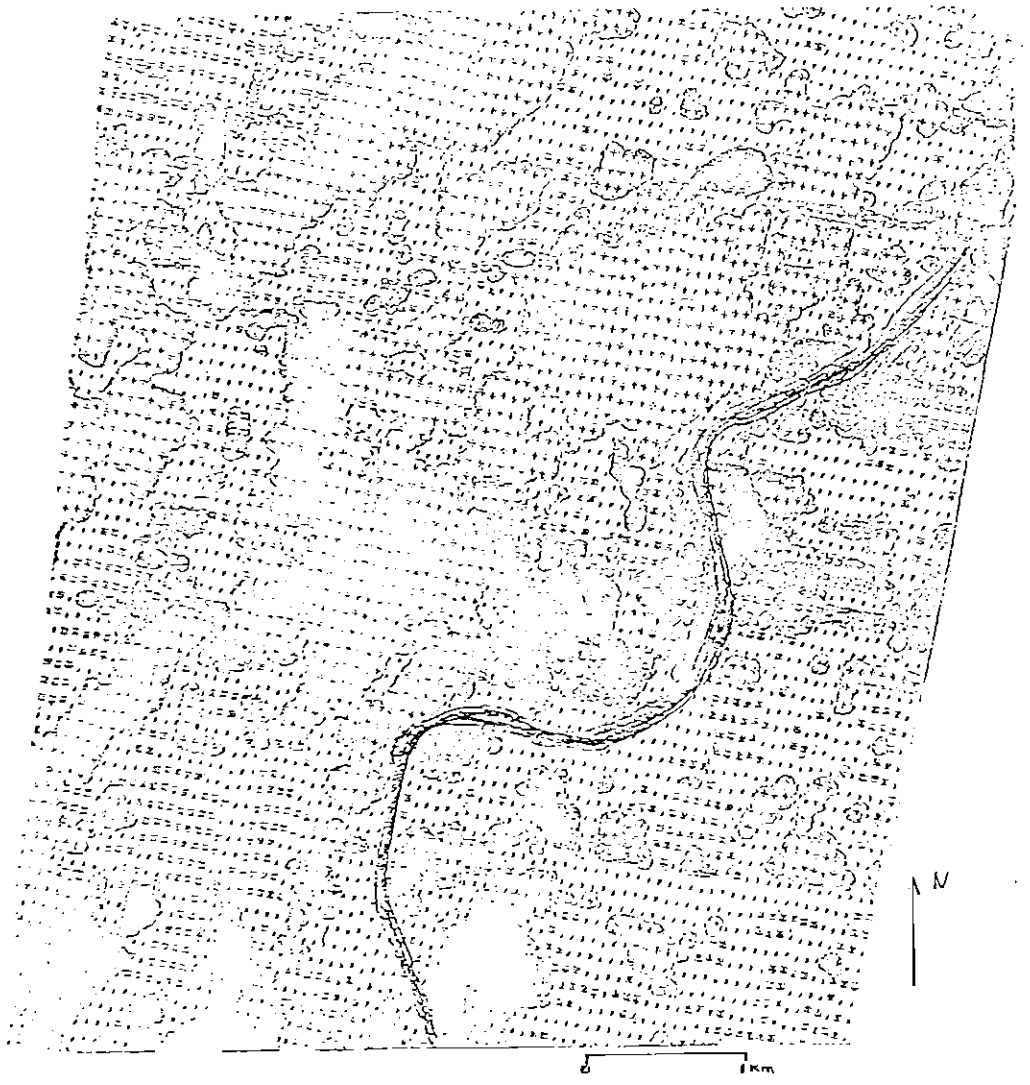

Figure 18. Classification of Douglas County Area.
Note the elongation of the plus cluster along the river.

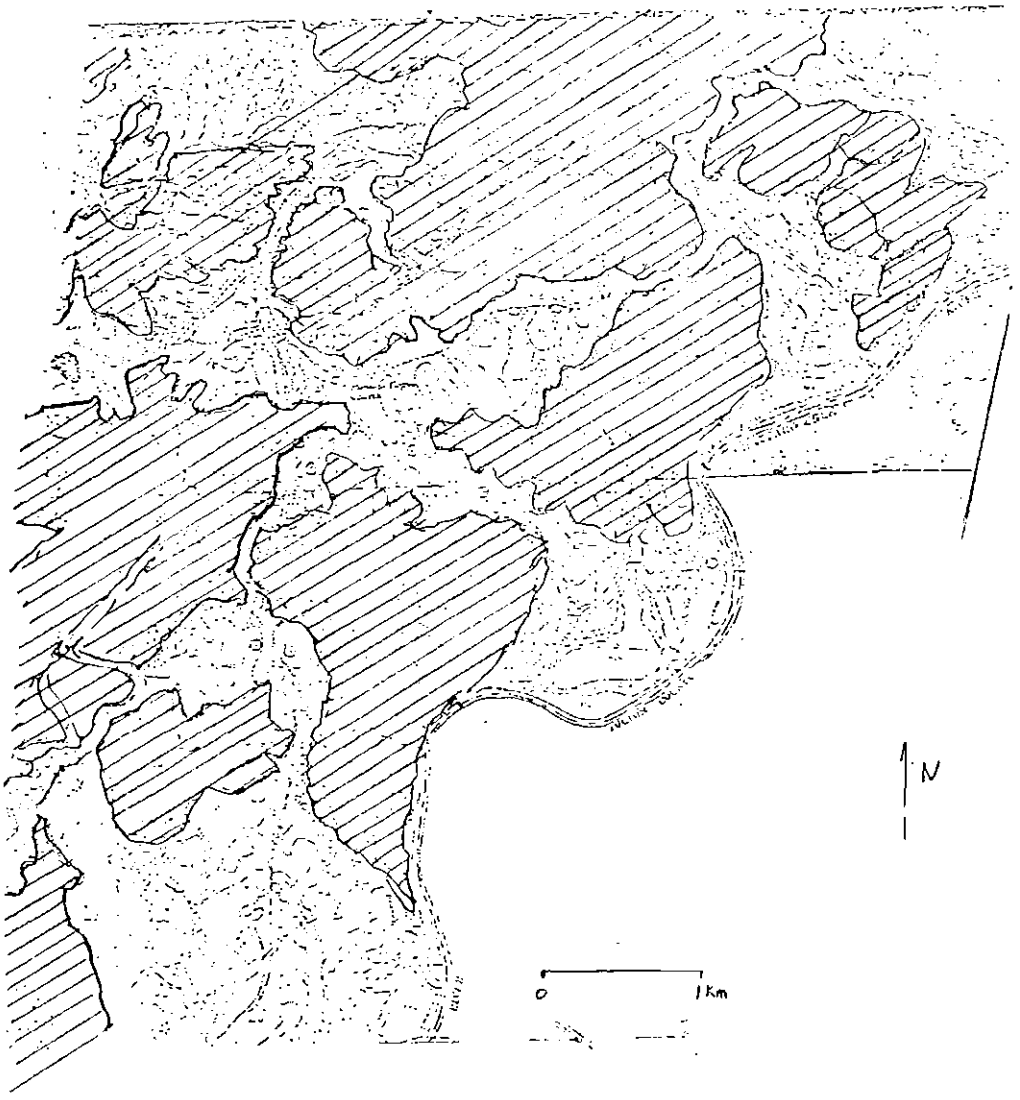

Figure 19. Soils Map of Douglas County Area.
Hatched Area is Loulsa Fine Sandy Loam.

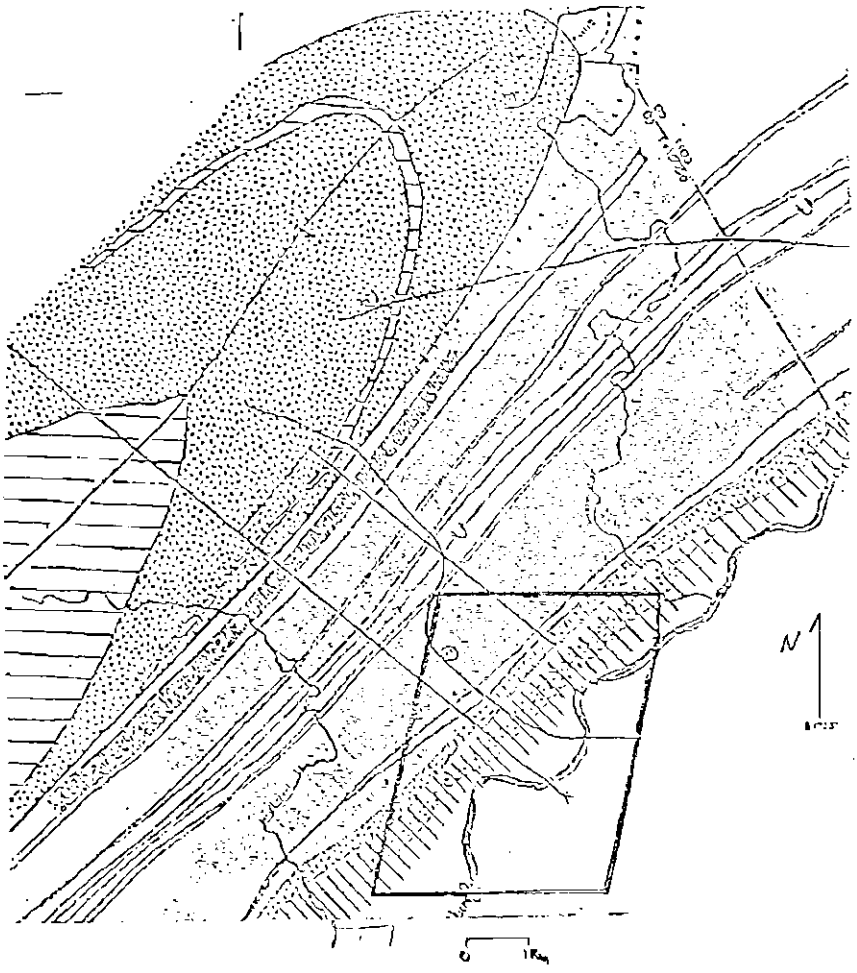

Figure 20. Geologic Map of Douglas County Area.
Boxed area is Clustered Area.

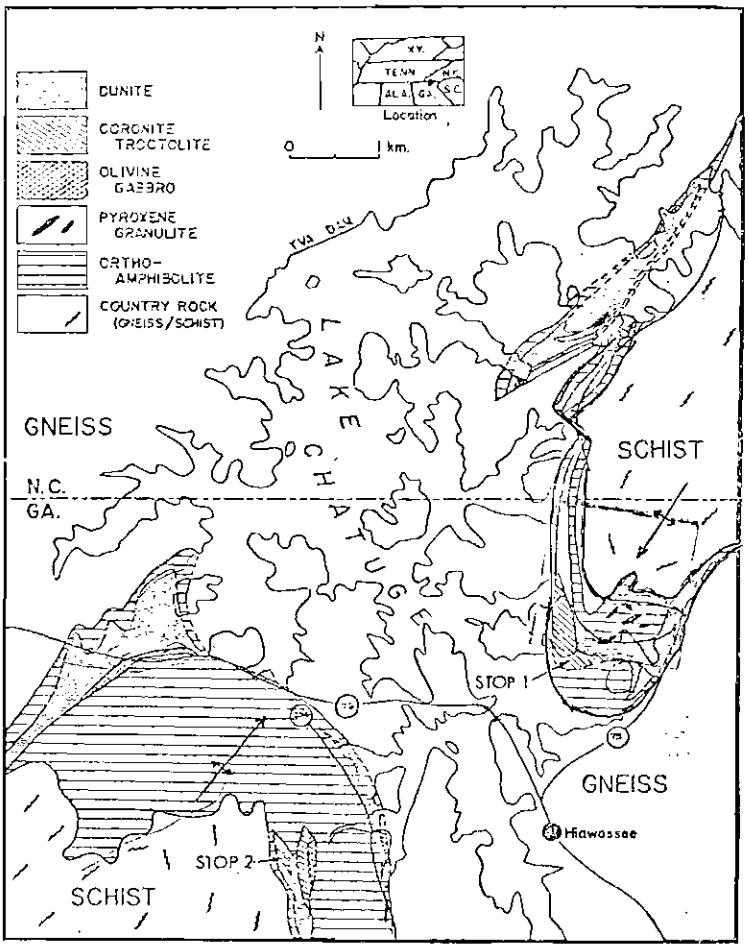

Figure 21. Geologic Map of Lake Chatuge Area.
Boxed Areas in Figures 21 and 22 Correspond to the Same Area.

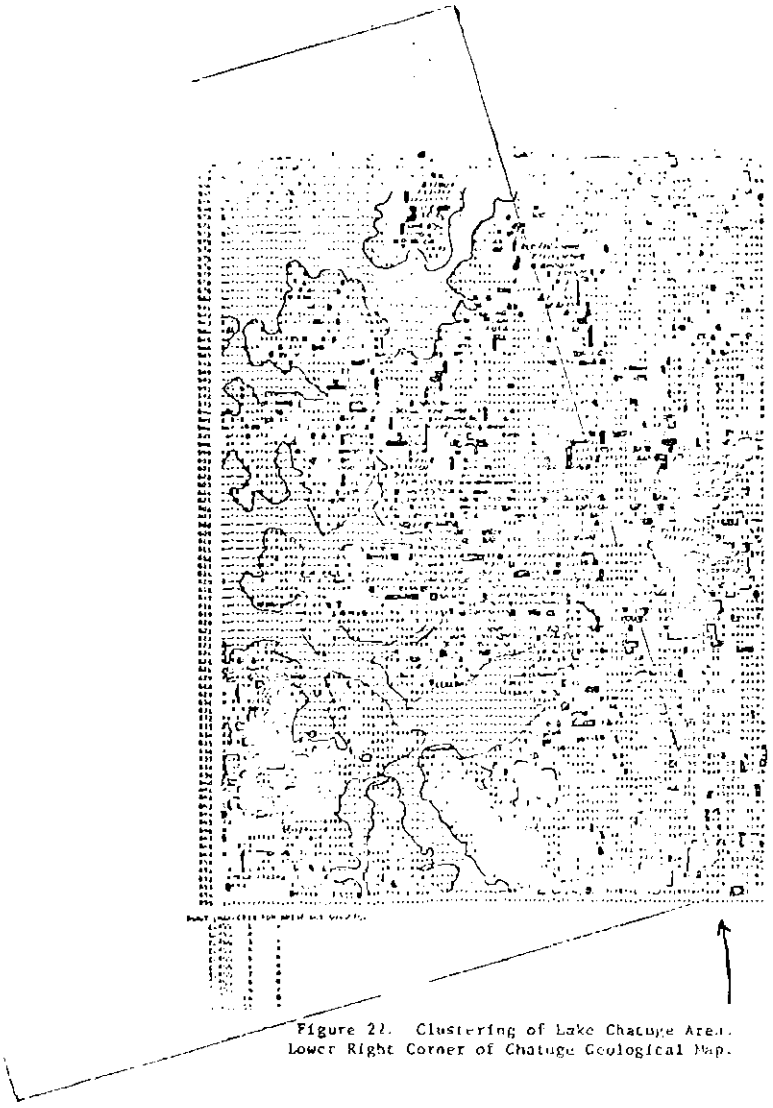

Figure 22. Clustering of Lake Chatuge Area. Lower Right Corner of Chatuge Geological Map.

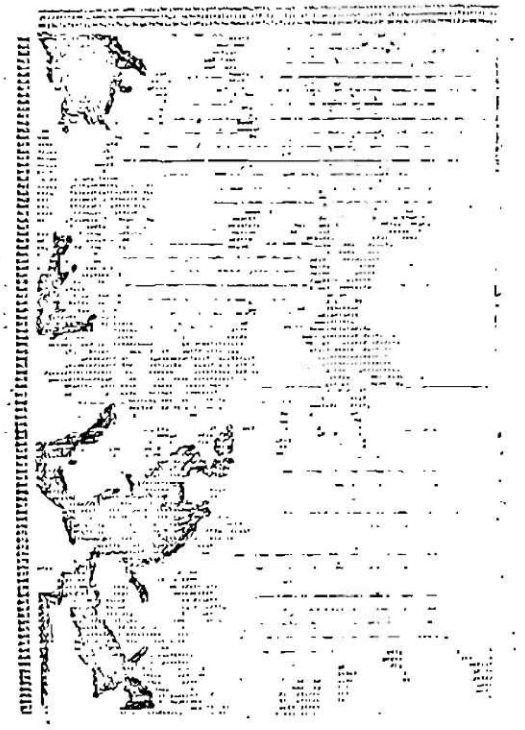

Figure 23. Unscaled Linear Classification of Lake Chatuge Area.

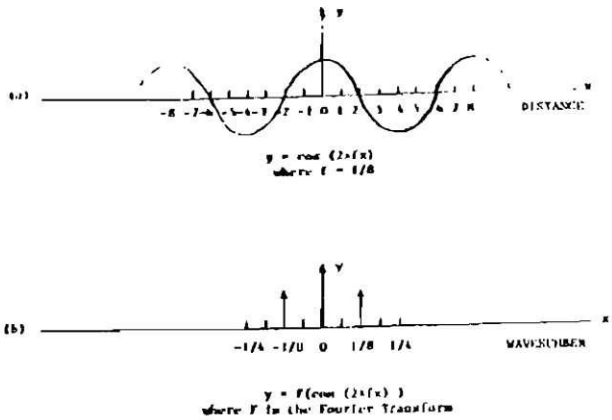

Figure 24. Pure Cosine Wave (a) and its Fourier Transform (b)

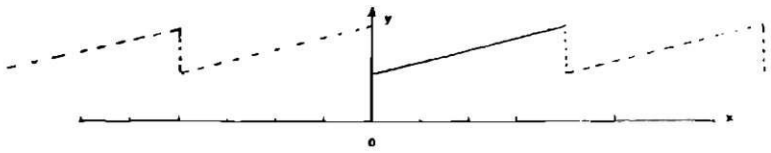

Figure 25. Ramp Function Between x = 0 and x = 4 - Cyclic Character Assumed by Fourier Transform is Dashed.

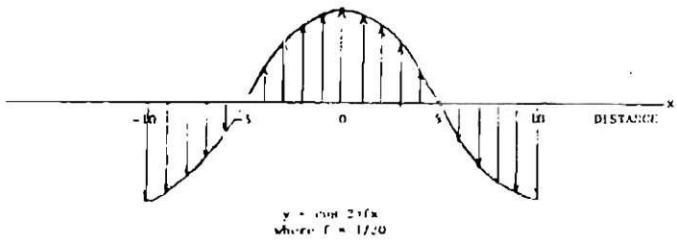

Figure 26. Adequate Sampling of Cosine Function

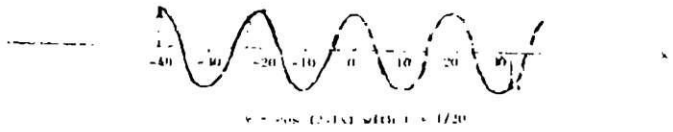

Figure 27. Inadequate Sampling Rate Gives Erroneous Digital Description of Function. (Dashed Line)

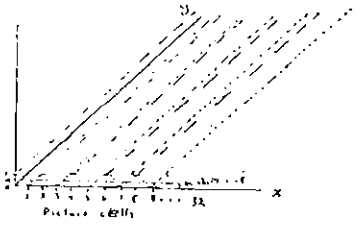

Figure 28. Sample Function #1.

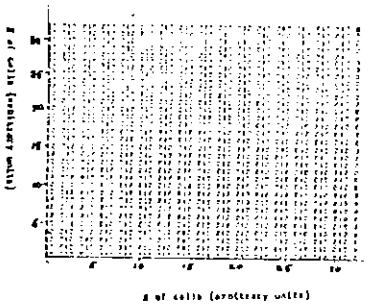

Figure 29. Digital Representation of Figure 5.

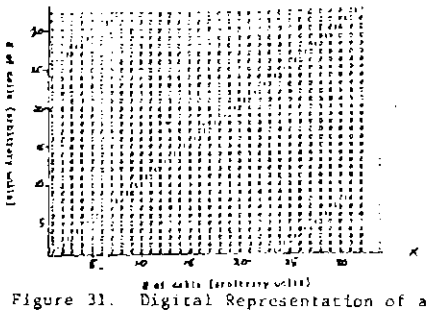

Figure 31. Digital Representation of a of a Noncyclic Linear Trend.

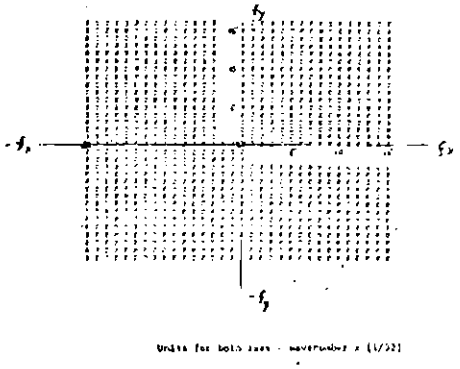

Figure 30. Star Diagram of the FFT of Figure 6.

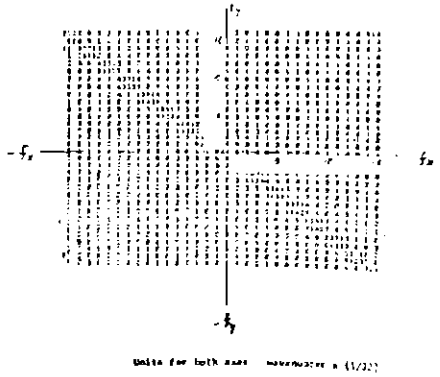

Figure 32. Star Diagram of the FFT of Figure 8.

Figure 33. FFT of Channel 7 of LANDSAT Data. Cumberland Mall Area.

Figure 34. FFT of Channel 4 of LANDSAT Data. Cumberland Mall Area.

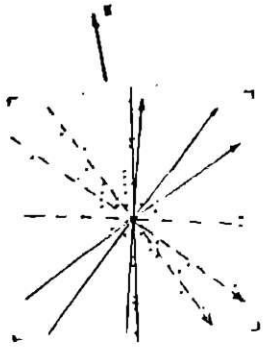

Figure 35. FFT of Channel 7 of LANDSAT Data. Carters, Georgia.

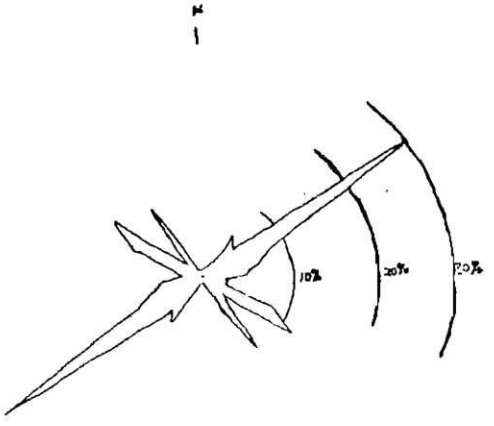

Figure 36. Geologic Joing Rose Diagram Carters, Georgia.

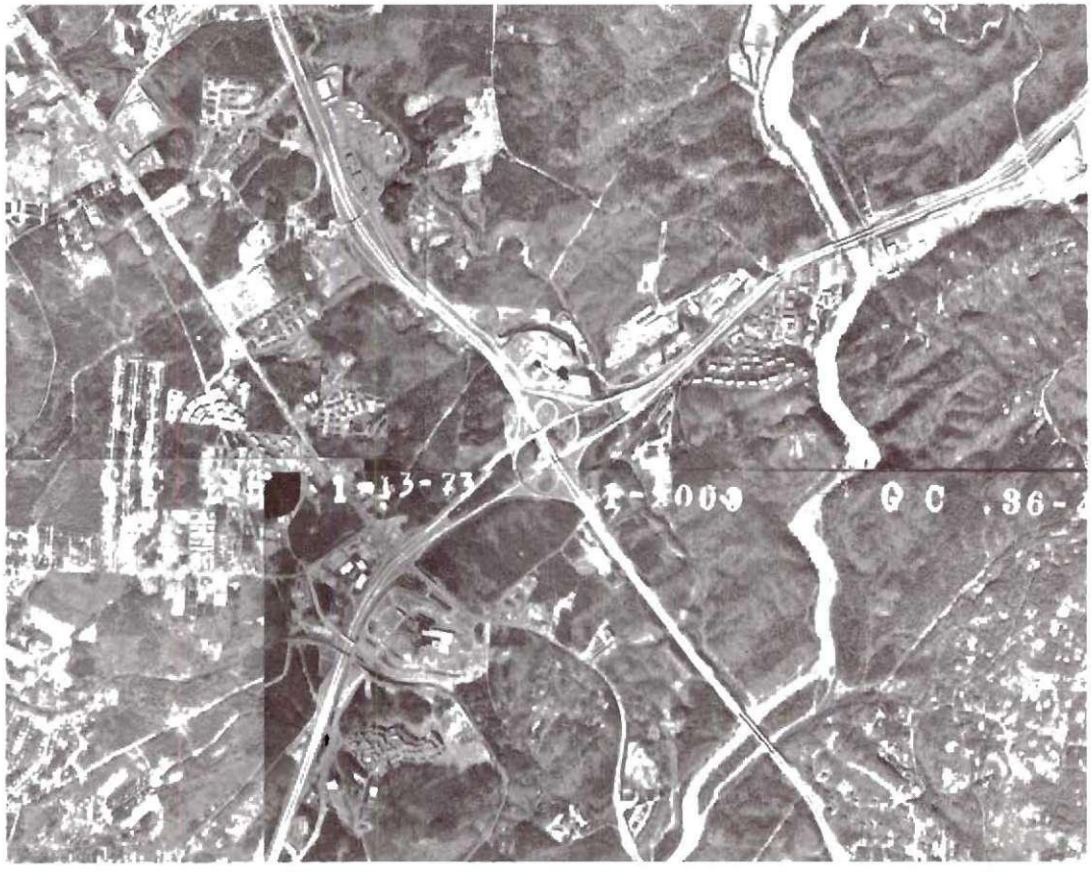

FIGURE 37.- AERIAL PHOTOGRAPH OF CUMBERLAND MALL AREA

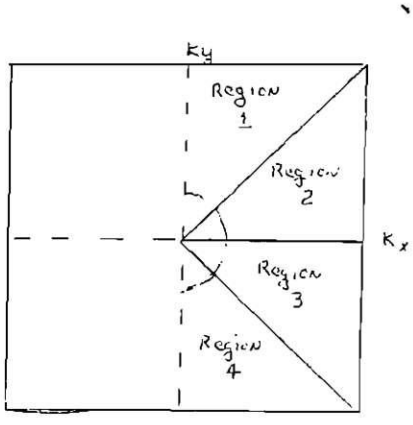

FIGURE 38.- FOUR ANGULAR REGIONS OF THE FFT STAR DIAGRAM USED TO EMPHASIZE LINEAR TRENDS

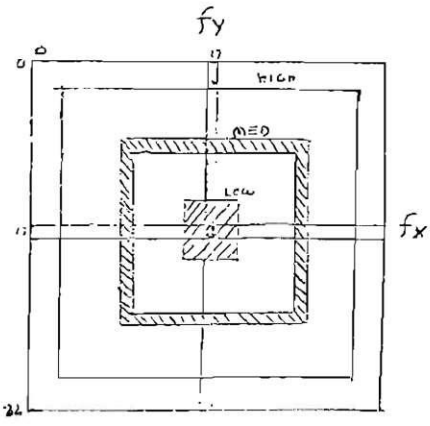

FIGURE 39.- THREE REGIONS OF THE FFT STAR DIAGRAM USED TO DIFFERENTIATE BETWEEN HIGH, MEDIUM, AND LOW FREQUENCY LANDSAT DATA PATTERNS

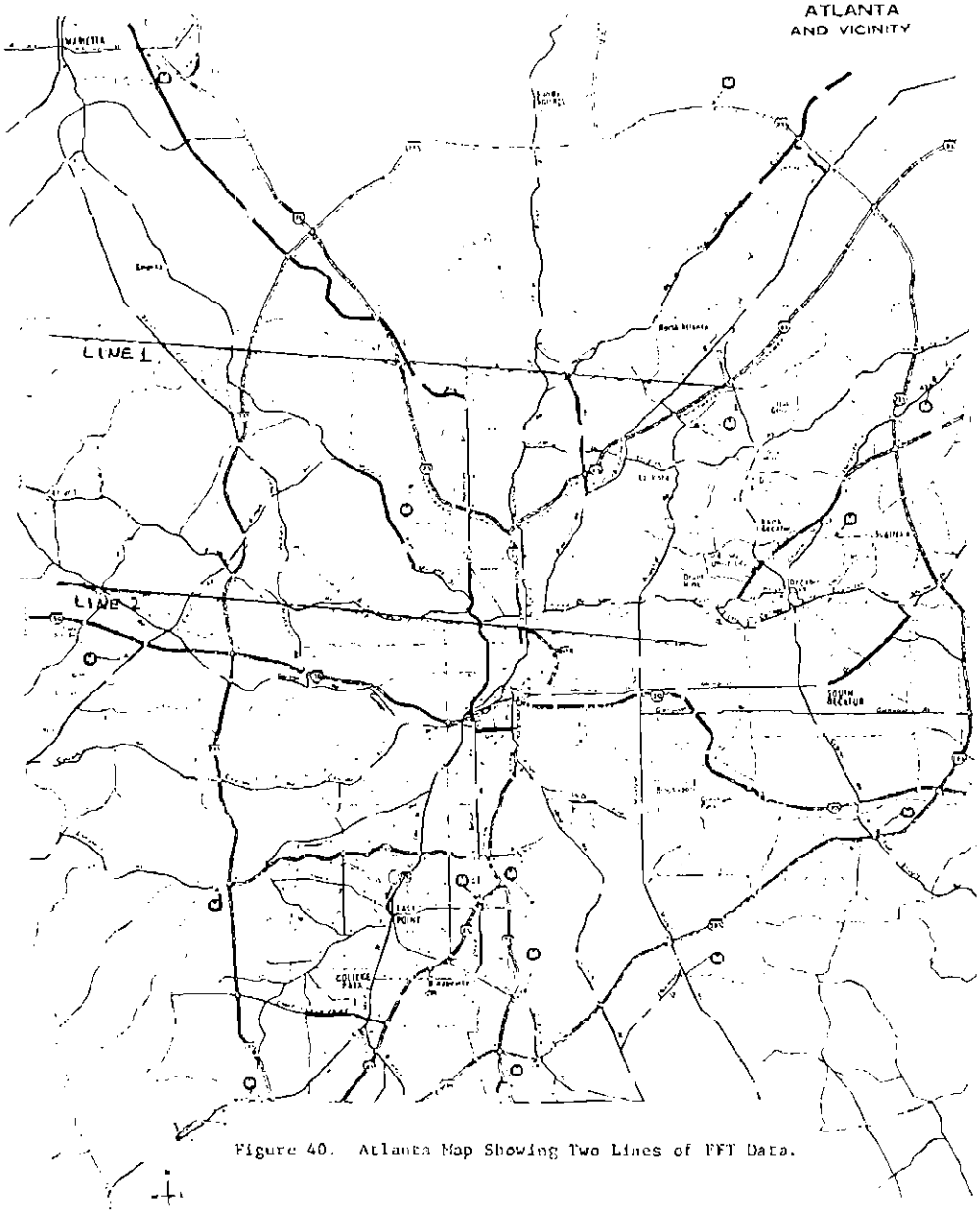

Figure 40. Atlanta Map Showing Two Lines of FFT Data.

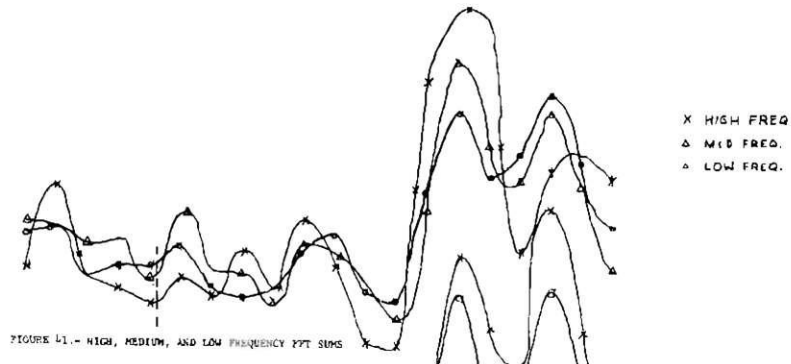

FIGURE 41.- HIGH, MEDIUM, AND LOW FREQUENCY FFT SUMS

ATLANTA - LINE 1

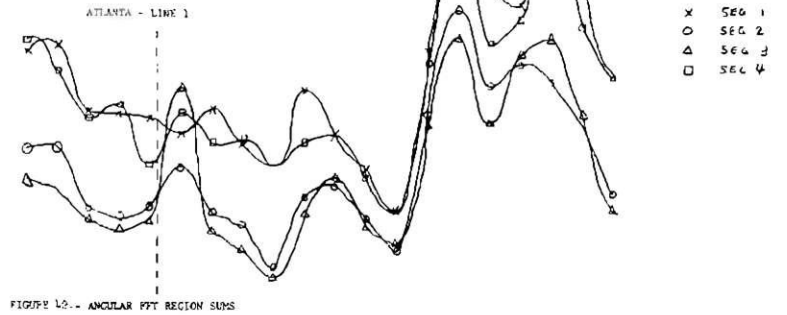

FIGURE 42.- ANGULAR FFT REGION SUMS

ATLANTA - LINE 1

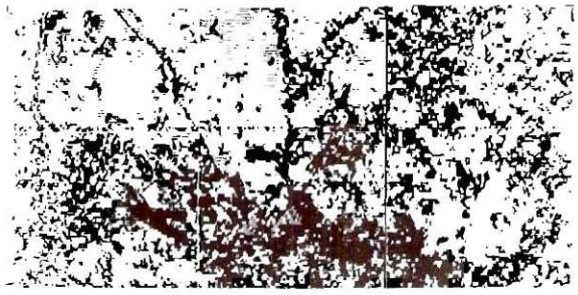

FIGURE 43.- COMPUTER MAP OF ATLANTA AREA

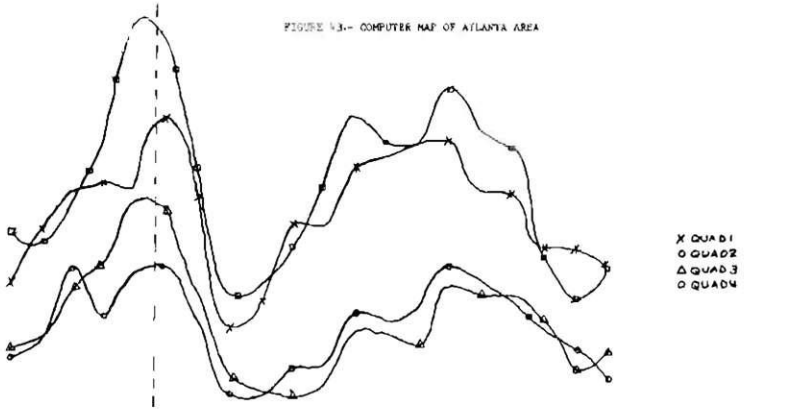

FIGURE 44.- ANGULAR REGION FFT SUMS

ATLANTA - LINE 2

REFERENCES

1. Baumgardner, M. F., et al., "Identification and Mapping of Soils, Vegetation, and Water Resources by Computer Analysis of ERTS MSS Data," NASA-CR 132956 (1973).

2. Bergland, G. D., "A Guided Tour of the Fast Fourier Transform," IEEE Spectrum, July 1969, 41-53.

3. Bond, A. D. and Atkinson, R. J., "An Integrated Feature Selection and Supervised Learning Scheme for Fast Computer Classification of MSS Data," Proc. of Conference on Earth Resources Observations and Information Analysis Systems, Tullahoma, Tenn., (March 1972).

4. Cooley, J. W., and Tukey, J. W., "An Algorithm for the Machine Calculation of Complex Fourier Series," Mathematics of Computers, 19, 297-301 (April 1965).

5. Crane, R. B. and Richardson, W., "Rapid Processing of Multispectral Scanner Data Using Linear Techniques," Proc. of Conference on Earth Resources Observations and Information Analysis Systems, Tullahoma, Tennessee (March 1972).

6. Crawford, T. J. and Medlin, J. H., "Brevard Fault Zone in Western Georgia and Eastern Alabama," Georgia Geological Survey Guidebook 12 (1974).

7. Duda, R. O., and Hart, P. E., Pattern Classification and Scene Analysis, John Wylie and Sons, New York (1973).

8. Ellefson, R., Swain, P., and Wray, J., "Urban Land-Use Mapping By Machine Processing of ERTS I Multispectral Data - A San Francisco Bay Area Example," Proc. of Conference on Machine Processing of Remote Sensing Data (1973).

9. Gramenopoulos, Nicholas, "Finding the Earth in ERTS," Optical Spectra, 26 (August 1974).

10. Haralick, R., "Adaptive Pattern Recognition of Agriculture in Western Kansas by Using a Predictive Model in Construction of Similarity Sets," Proc. of Symposium on Remote Sensing of Environment, (September 1968).

11. Hartley, M. E. III, and Penley, H. M., "The Lake Chatuge Sill Outlining the Brasstown Anteform," Georgia Geological Survey Guidebook 13 (1974).

12. Hayes, W. L., and Winkler, R. L., Statistics - Probability, Inference, and Decision, Holt, Rinehart, and Winston, Inc., New York (1971)

13. Hermann, L. A., "Geology of the Stone Mountain-Lithonia District, Georgia," Ga. Geol. Survey Bull. #61 (1954).

14. Ho, Y. C. and Agrawala, A., "On Pattern Classification Algorithms: Introduction and Survey," Proc. IEEE, 56, 2101-2114 (December 1968).

15. Langrebe, D. A., "Automatic Processing of Earth Resources Data," Proc. of Second Annual Aircraft Program Status Review (1969).

16. Nagy, G., "State of the Art in Pattern Recognition," Proc. IEEE, 56, 836-963 (May 1968).

17. Saunders, D. F., Thomas, G. E., and Kinsman, F. E., "ERTS Imagery Promising as Petroleum Exploration Tool," World Oil, 53 (March 1974).

18. Shackelford, R. G. and Walsh, J. R. Jr., "A Study to Analyze Six Band Multispectral Images and Fabricate a Fourier Transform Detector," Georgia Tech EES Final Report #A-1592 (June 1965).

19. Smedes, H. W., "Geologic Studies of Yellowstone National Park Using an Electronic Enhancement System," U. S. Geological Interagency Report 181 (1972).

20. Spann, G. W. and Faust, N. L., "Study of USDA/NASA Land Use Classification System," Georgia Tech EES Interim Technical Report Project A-1621 (December 1974).

21. Su, M. Y., Jayroe, R. R., and Cummings, R. E., "Unsupervised Classification of Earth Resources Data," Proc. of Conference on Earth Resources Observation and Information Analysis Systems, Tullahoma, Tenn. (1972).

22. Ulaby, F. T., and McNaughton, J., "Classification of Physiography from ERTS Imagery," Photogrammetric Engineering and Remote Sensing, XLI, No. 8 (August 1975).

23. Wells, R. D., et al., "Soil Survey of Douglas County, Georgia," USDA Soil Conservation Service, Series 1959, No. 1 (November 1961).

24. Westin, F. C., "Remote Sensing of Physiographic Soil Units of Bennet County, South Dakota," NASA-CR-131229 (1973).

25. Wylie A. et al., "Users Guide and Software Documentation for the Algorithm Simulation Test and Evaluation Program (ASTEP), Johnson Spacecraft Center Internal Note #73-FM-71 (December 1974).